

R E P O R T R E S U M E S

ED 003 884

48

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION--IV. PROGRESS REPORT NO. 4.

BY- LANE, HARLAN AND OTHERS

MICHIGAN UNIV., ANN ARBOR, COLL. OF LIT., SCI., ARTS

REPORT NUMBER NDEA-VI-78-4

PUB DATE NOV 63

REPORT NUMBER PR-4

REPORT NUMBER D5613-1-P

CONTRACT OEC-3-14-013

EDRS PRICE MF-\$0.36 HC-\$8.32 208P.

DESCRIPTORS- *SPEECH, *PERCEPTUAL DEVELOPMENT, *LANGUAGE RESEARCH, *PROGRAMED INSTRUCTION, EXPERIMENTAL PROGRAMS, THEORIES, LINGUISTICS, *SPEECH INSTRUCTION, ANN ARBOR, MICHIGAN

THE PROGRESS OF AN EXPERIMENTAL RESEARCH PROGRAM IN SPEECH CONTROL WAS REPORTED. DISCUSSION TOPICS WERE (1) PROGRAMED LEARNING OF A SECOND LANGUAGE, (2) A CRITICAL REVIEW OF THE MOTOR THEORY OF SPEECH PERCEPTION, (3) VISUAL CONTROL OF PHONEMIC CONTRASTS (A TEST OF THE MOTOR THEORY OF SPEECH PERCEPTION), (4) SINGLE STIMULUS PROPERTIES IN THE IDENTIFICATION OF COMPLEX TONES, AND (5) A TECHNIQUE FOR PRODUCING KNOWN SOUND PRESSURES AT THE EARDRUM. (RELATED REPORTS ARE ED 003 883-ED 003 887.) (RS)

ED003884

PLEASE DO NOT REMOVE AND
RETURN TO USOE

78
No 4
C5613-1-P

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS

DEPARTMENT OF PSYCHOLOGY
BEHAVIOR ANALYSIS LABORATORY

Progress Report No. 4
April 1, 1963, to November 1, 1963

Experimental Analysis of the Control of Speech Production and Perception: IV

Under contract with:

Language Development Section
U. S. Office of Education
Contract No. OE-3-14-013
Washington, D. C.

Administered through:

November 1963

OFFICE OF RESEARCH ADMINISTRATION • ANN ARBOR

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS

Department of Psychology
Behavior Analysis Laboratory

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the
person or organization originating it. Points of view or opinions
stated do not necessarily represent official Office of Education
position or policy.

Progress Report No. 4

April 1, 1963, to November 1, 1963

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION: IV

ORA Project 05613

under contract with:

LANGUAGE DEVELOPMENT SECTION
U. S. OFFICE OF EDUCATION
CONTRACT NO. OE-3-14-013
WASHINGTON, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

November 1963

CONTENTS

PROGRAMMED LEARNING OF A SECOND LANGUAGE

THE MOTOR THEORY OF SPEECH PERCEPTION: A CRITICAL REVIEW

VISUAL CONTROL OF PHONEMIC CONTRASTS: A TEST OF THE MOTOR THEORY
OF SPEECH PERCEPTION

ATTENTION TO SINGLE STIMULUS PROPERTIES IN THE IDENTIFICATION OF
COMPLEX TONES

A TECHNIQUE FOR PRODUCING KNOWN SOUND PRESSURES AT THE EARDRUM

APPENDIX

CURRENT STAFF

LIST OF PUBLICATIONS

PROGRAMMED LEARNING OF A SECOND LANGUAGE

**Harlan Lane
The University of Michigan**

Programmed learning of a second language

Harlan Lane

The University of Michigan¹

Introduction

The cardinal requirement of programming, "specify the terminal behaviors," has nowhere had a more salutary effect on educational technology than in the area of second-language learning. Although language pedagogy turned to linguistic science for these specifications over two decades ago, the later arrival of programmed instruction emphasized, and placed a new perspective on, the contribution of descriptive and structural linguistics. The new emphasis is that linguistic specifications are a sine qua non for the conditioning enterprise; they codify group norms for verbal behavior in the relevant language community and, therefore, provide the programmer with criteria for acceptable forms and sequences of responses. The new perspective is that a linguistic analysis of terminal behaviors is not a prescription for second-language learning; this is the task of the programmer.

The good fortune of the language programmer (and student) in finding an entire discipline devoted to the systematic specification of terminal behaviors is overreached by an even greater fortune: the techniques of the laboratory analysis and control of behavior, upon which programming is based, are more literally applicable to the conditioning of second-language behavior than to most of the other repertoires that programming has con-

fronted. The objectives of most programs are restricted to changing the written, verbal behavior of a student in response to textual stimuli. The prerequisite motor responses and discriminations among orthographic stimuli have already been conditioned and are taken for granted. The behaviors involved in reading and understanding English and in writing at least a "reasonable" answer are also assumed. As Gilbert (1962) has put it: "Our student is not a master of the subject because he cannot make the mastery responses on the right occasions, not because he is unable to make those responses at all... The responses of mastery are there...." This type of verbal conditioning, involving a restructuring of the student's highly articulate first-language repertoires, has a limited resemblance to laboratory research on the control of behavior. Transitional behavior effected by shifting contingencies of reinforcement, patterns of discriminative stimuli, or both, is little studied and poorly understood compared to the shaping, maintenance, and extinction of behavior with initially naive organisms.

The programming task of restructuring an extant verbal repertoire may be contrasted to that of conditioning second-language fluency. In the latter case, there is nothing extrapolative in the application of laboratory techniques and nothing metaphorical in the use of concepts gained from a functional analysis of behavior in the laboratory. New discriminations (auditory, textual) must be conditioned, new forms of response (vocal, written) must be differentiated, concurrent responses of differing topography must each be brought under the control of appropriate discriminative stimuli, novel chains of topographically novel responses must be established, and so forth. These

behaviors must be conditioned largely de novo, although it is never possible, as we shall see, to overlook sources of interference and enhancement from the first-language repertoires.

Two other developments augur well for the success of second-language programming. The interest of the communication sciences in verbal, especially vocal, behavior has led to a greater sophistication and instrumental capability in describing and analyzing the vocal response than is available for any other behavior (Fant, 1958). These disciplines have also contributed significant techniques and data toward an understanding of the listener's behavior as well as that of the speaker (Licklider and Miller, 1951). Another source of methods and findings for the control of verbal behavior is less well developed but bound to have equal or greater impact on language pedagogy; this is the rapidly growing discipline of psycholinguistics (Saporta, 1961).

If second-language programming is aided by these four "sources of strength," it also has four countervailing antecedents in language pedagogy which must be set aside: an unsuitable model of learning, undue emphasis on language aptitude, subjective evaluation of student performance, and incompetent classroom research.

The language teacher has operated frankly with a sunburn model of learning (Lane, 1962c) and her techniques are as inefficient as this characterization of learning is inappropriate. The teacher, prime source of knowledge, light (and, occasionally, heat) "exposes" students to the language and its principles. She is aided in this endeavor by the language laboratory, now a commonplace in secondary education (Morton, 1961b), which

also exposes the students to the material in limited doses; 20 minutes is generally considered the maximum safe exposure in one treatment. When classroom practices permit active participation by the student at all, it is usually in the form of repetitive drills which are presumed to "fix" or "stamp in" the material through repetition alone. The "brighter" students "soak up" the material and become "enlightened." Dull students, who fail to learn, are simply not "sensitive" or "receptive."

The remarkable plasticity of behavior is ignored by this model. Instead of acknowledging that the inadequacies of the students are the product of inadequate technique, and then spending all effort to improve technique, the key role in language learning is assigned to ability or endowment, thus placing the process beyond the teacher's control—and responsibility. The construct "language aptitude" is the deus ex machina that spares the language-learning story its tragic ending. Psychologists have not helped to refocus attention on the critical variables, the manipulable variables, in language-learning by constructing so-called aptitude tests. Indeed, since it is invariably true that the student who takes the test is assigned an aptitude score, the presence of the score lends credence to the reality of aptitudes. This gambit has as much practical effect as lifting oneself by one's own bootstraps, and leaves the language teacher in about the stance implied. The naive assumption that aptitude tests measure innate capacity is no longer acceptable. As one noted student of individual differences points out, "It should be obvious that all psychological tests measure the individual's current behavior, which inevitably reflects the influence of prior learning" (Anastasi, 1961). The task of the language teacher or programmer is to

destroy the test—retest reliability of language aptitude tests.

Evaluations of pedagogical techniques in second-language learning are most often grossly multi-dimensional, idiosyncratic and subjective. "The objectives of foreign language study still include such goals as understanding the foreign culture, appreciating the great works of literature in the language, increasing English vocabulary, learning about grammar—Latin even taught us how to think!" (Sapon, 1963). The merit of multiple and complex objectives such as these is that the inadequacy of language pedagogy in attaining them cannot be assessed; thus, "achievement" in these areas may be offered as mitigating the limited effect of the teacher in conditioning the language behaviors themselves. Even when the objectives are somewhat more circumscribed as : "a spoken command of the language," the criteria for an acceptable response include such notions as phonetic accuracy, response latency, intonational accuracy, appropriateness of vocabulary, richness of vocabulary, use of idioms, propriety of style, etc. Only rarely has the criterion behavior been adequately specified, and even under those circumstances, the assessment device usually involves a panel of judges and is heir to the traditional problems of inter- and intra-judge reliability and validity.

Prior research on second-language learning² has been, with few exceptions, seriously inadequate. Carroll (1962a) seems to damn it all, in an extensive survey of research in this area, by concluding with the faint praise: "it is clearly within the realm of possibility that the results of such research could make language teaching more effective and efficient."

The inadequacies of the research are several. "Whereas psychologists perform experiments with insufficient relevance to foreign language teaching, members of the foreign language teaching profession perform studies with insufficient experimental rigor" (Carroll, 1962a). Even when relevance and rigor have prevailed, studies in this area have suffered the many evils of comparison-group designs (vide Lane, 1961b). Furthermore, the comparison is usually among the effects of complex experimental treatments, many-faceted and extending over many months, evaluated all too often either anecdotally or by tests of doubtful validity and reliability. These properties of the experimental method are in part the result of competing pedagogical and research objectives in the classroom setting. As a consequence, the findings of even the most carefully executed studies have little generality. No wonder that Carroll (1960) finds no research basis for educational policy on foreign language teaching: "However sympathetic we may be to [current] recommendations, they are often based on assumptions about which the most charitable thing to say is that they have not been proven."

The character of current second-language programming reflects these several antecedents. Many programs draw extensively from the findings of linguistics, both in specifying the terminal behaviors and in determining sequencing based on a contrastive analysis of the student's first and second languages. A few have systematically applied laboratory techniques for the control of behavior. A lesser number reveal the impress of findings in the communication sciences, especially in acoustic phonetics; psycholinguistics has had little impact as yet. Finally, many programming efforts have inherited the tradition of educational research in language pedagogy. The

program is admixed with arbitrary amounts of classroom and language-laboratory instruction and drill and the effects of the complex enterprise are compared in cursory fashion to those of more traditional methods. The failure to provide a detailed account of the terminal repertoires actually attained by students who have completed the second-language program is almost universal.

The present discussion of the basic research underpinnings and applied findings in second-language programming proceeds in terms of a functional analysis of verbal behavior (Skinner, 1957). The discussion is behavioral, because the objective of second-language programming is to condition certain terminal behaviors (largely spoken and written); the discussion draws heavily on a functional analysis, because it is this analysis which describes the relations between the antecedents and consequences of behavior and the behavior itself; we must know these controlling relations if we are to arrange effective and efficient second-language learning.

A functional analysis of the terminal behaviors for a second-language program will usefully distinguish between formal repertoires, in which there is a point-to-point correspondence between stimulus and response, and thematic repertoires in which responses are controlled by common sets of variables but without formal correspondences (Skinner, 1960). The student (or child) who has learned only to imitate, read, and copy in a new language has acquired purely formal repertoires; the topography of his behavior in each case shows a point-to-point correspondence to the form of the stimulus conditions. However, he is as yet unable to respond appropriately to the nature of the stimulus conditions and this lack of thematic repertoires leaves him unprepared to join the verbal community. In sum, he knows how

to say things but he does not know what to say. Talking birds, mimics, typists, and stenographers may earn their keep by purely formal repertoires, but most adult verbal behavior reflects both formal and thematic sources of strength. Thus, the terminal behavior of the student will be determined both phonetically (or orthographically) by the audience and thematically by the environmental conditions.

The distinction between the two repertoires may be sharpened by noting that a minimal unit of behavior may be identified for a purely formal repertoire and, therefore, formal repertoires may be extended readily to new stimuli. The minimal unit may be the phoneme or the grapheme or it may be even more fragmentary. Once these stimulus-response units are made available through the conditioning of a formal repertoire, the student can respond adequately to novel stimuli by responding piecemeal—that is, with a series of unit responses never before arranged in that order. For example, minimal command of the formal repertoire involved in reading English text permits a passable attempt at, say, "marmoset," but nothing in the thematic repertoires that the student may have acquired previously enables him to say "marmoset" in the presence of marmosets, to make statements about marmosets, and so forth. The ability of formal repertoires to be extended to new stimuli is predicated on the availability of minimal units which require, in turn, the point-to-point correspondence of stimulus and response that characterizes these repertoires.

Formal Repertoires

We may subdivide the formal repertoires involved in second-language learning according to whether the stimulus and the response are each written or spoken. The distinction is important not only because the two classes of stimuli have different modalities and the two classes of responses have different topographies, but also because the four possible functional relations have different dynamic properties. When both stimulus and response are spoken, we term the behavior echoic; the name implies the point-to-point acoustic correspondence which characterizes this repertoire. In transcription, both stimulus and response are written. In textual behavior, a spoken response is controlled by a written stimulus. In dictation, the stimulus is spoken, the response written. In each of the first two formal classes, the stimulus and the response have similar form, acoustic or visual, but in all four classes it is possible to predict the response given the stimulus or to infer the nature of the stimulus if the response is known; this is another way of stating the point-to-point correspondence between stimulus and response that defines a formal repertoire.

Each of the four formal repertoires obviously requires that the student make different responses to physically different stimuli; it is perhaps less obvious that he must more often make the same response to different stimuli; different fonts of type typically evoke the same textual behavior, differing rates of speaking the same transcriptions, and so forth. When the properties of responding change following a change in the stimulus, we speak of stimulus discrimination; to the degree that responding remains invariant under this

change, we speak of stimulus generalization. The first requirement of a formal repertoire, then, concerns stimulus control: there must be generalization within arbitrarily defined classes of stimuli and discriminations between these classes. The second requirement concerns the properties of responding: response topography must show small, systematic variation within arbitrarily defined classes and large and abrupt changes from one class to the next. The third requirement of a formal repertoire concerns the coordination of changes in the stimulus with changes in the response. To take an example from a textual repertoire in English, in order to read the pair of words "pin pan," these requirements must be met: (1) the orthographic symbol "i" must be visually discriminated from "a," etc., but generalized with "i," "I," and that letter in various type fonts; similar requirements apply to "a" and to the other letters. (2) response topography must include distinct response classes such as $[p^h]$, $[I]$, and $[x]$ but there is also a range of permissible and necessary variation within each class; thus, the reading of "pin" may range from $[p^hIn]$ to $[p^hI:n]$ or to $[pIn]$. (3) $[p^h]$ must occur in the presence of initial "p," $[I]$ in the presence of "i," and so forth. Even if the letters "i" and "a" are discriminated and the responses $[I]$ and $[x]$ are differentiated, without the coordination required the textual behavior would be incorrect (e.g., $[p^hxn - p^hIn]$).

Before considering each of the four formal repertoires, the sections that follow describe some recent experiments on discrimination, differentiation, and coordination. Laboratory research on these behavioral processes provides methods and findings that may be used to arrange their acquisition in second-language programming.

Discrimination learning and transfer

The acquisition and maintenance of stimulus discriminations have been studied extensively with human and infra-human organisms in the psychological laboratory. These experiments involve most often stimuli that are synthesized in the laboratory and varied along a single dimension, such as pure tone frequency. Controlled experiments on discrimination learning with synthetic stimuli varying in several dimensions, or with stimuli from "natural language" (spoken or written), are rare. A description of four representative experiments will illustrate methods and findings that are relevant to the programming of second-language discriminations involved in the formal repertoires.

In a series of experiments on discrimination learning with speech stimuli in a second language, Lane (1961) has observed extremely rapid acquisition of differential stimulus control, suggesting the transfer of the subject's first-language discriminations. When a Spanish phoneme was presented as S^D (a stimulus correlated with reinforcement for responding) and a variety of English approximations were presented as S^A s (correlated with nonreinforcement), the subjects stopped responding to S^A s after the first few presentations. Indeed, most of the errors in learning were due to "over-discriminating," that is, failing to respond to the allophonic variants of S^D throughout the sequence of stimuli. In one study, a self-instruction program was prepared to teach discriminations among the vowels and consonants of Spanish. The program was subdivided into 14 frames, each of which included allophones of a particular Spanish phoneme (S^D) and those of other English and Spanish phonemes (S^A 's). The 14 frames as well as the 60 stimuli within each frame were sequenced accord-

ing to a tentative schedule of difficulty for the English-speaking student. When the subject responded (button-press) to an S^D or failed to respond to an S^A during the 4-sec interstimulus interval, a point was added to a counter in front of him. Failing to respond to S^D , or a response to an S^A , cost him a point. Figure 1 shows that most of the frames in this program were mastered

 Insert Fig. 1 about here.

to a criterion of eight or less errors in a single trial. These data and others reported suggest that discrimination learning among the sounds of a second-language, at least one within the same family as the subject's native language, may be expected to proceed quite rapidly when conditions are optimized because the subject transfers first-language discriminations to the new task. Additional support for this conclusion was obtained when the program described above was administered without reinforcement both before and after programmed learning. For each of the 14 frames, Fig. 2 shows the number of

 Insert Fig. 2 about here.

errors made during programmed learning plotted against the percent increase in correct responses from the pre-test to the post-test. There are some frames in which the number of errors and the increase in percent correct are small; we may infer that this high baseline of discrimination is the product of first-language training. There are also several frames in which the number of errors is again small but the increase in percent correct is large. This reflects a discontinuity in the time-course of discrimination learning (a few

errors followed by an abrupt increase to near-perfect performance) which has been called "concept attainment" in other quarters; it also reflects the subject's prior discriminative history. The more gradual development of differential stimulus control that is normally associated with discriminative learning was shown by only a few frames, such as the one involving the trilled Spanish /r/ in intervocalic position as S^D and the English /r/ in that position as S^A .

Suppes et al. (1962) have performed several studies of discrimination among Russian phonemes and have examined the fit of mathematical models to the observed time-course of discrimination learning. In one study, a basic list of Russian syllables was constructed with 32 initial consonant phonemes and 5 vowels. The 144 syllables were grouped into contrasting pairs, differing only in the initial consonant. Twenty-six consonant-vowel (CV) contrasts were selected and each was presented in four permutations: $C_1V_1:C_2V_2$; $C_2V_1:C_1V_1$; $C_1V_2:C_2V_2$; and $C_2V_1:C_1V_2$. The list of 104 contrast pairs was recorded in random order by a linguist and presented for one-half hour daily, on five days, to individual subjects for judgments of same or different. Incorrect responses were followed by a flash of light. A contrastive analysis of English and Russian and some pilot studies suggested that certain contrast pairs would prove easier to discriminate than others; these were presented to one group of twenty subjects while those considered difficult were presented to another group of twenty.

The first group showed a drop in errors over the five days from 11 to 2%, the second from 22 to 10%. The low initial error rates (high baselines of discrimination) probably reflect transfer of first-language discriminations,

discussed above. Based on the error distributions obtained in this experiment, the authors present a rank ordering of the Russian consonants and vowels with respect to difficulty in discrimination learning for English-speaking subjects; the order accords well with that predicted from a contrastive analysis of the two languages.

In a subsequent study, Suppes, Crothers, and Weir (1962) examined discrimination learning among Russian vowel phonemes and the effect of sequencing stimulus presentations according to expected difficulty. The subject was presented with a recorded series of sets of five stimuli during six, daily 15-minute sessions distributed over a period of 11 days. Each stimulus was a Russian word containing one vowel preceded and/or followed by a consonant. The subject was instructed to select the alternate stimulus in each set which contained the same vowel as the first stimulus in the set; he was then told the correct answer. One group of subjects ($N = 44$) received a fixed series of items in random order (R), a second ($N = 54$) received progressively longer and more difficult items in each daily session (P).

Figure 3 shows the proportion of errors on each experimental day for the

 Insert Fig. 3 about here.

subjects in each group. On day 1, the programmed subgroups naturally show fewer errors since they were presented with only the easier discriminations in the fixed series presented to group R. On days 5 and 6, when both groups were presented the same discrimination series, the programmed group committed a reliably greater number of errors. In interpreting this finding it must be

remembered that the programmed group had less training on certain of the difficult discriminations because of the sequencing of stimuli presented to them. The procedure of vowel-phoneme matching from a set of alternatives permits the preparation of a confusion matrix which can serve as a guide to sequencing and repetition in programming the acquisition of the relevant discriminations (cf. Sapon and Carroll, 1958, and discussion below).

These three studies and others (Pimsleur, 1963; Lane & Moore, 1962; Lane, 1962^a) illustrate the extensive transfer of first-language discriminations in second-language learning, whether the task be identification, detection of a difference, or recognition, respectively (see Gallanter, 1962). The second experiment by Suppes et al. and that by Swets et al., to be described, show, however, that rapid and relatively error-free discrimination learning cannot be assured solely on the basis of an a priori sequencing of item difficulty. Among other requirements, there must be an iterative process of program revision, including the revision, addition and deletion of frames and sequences, based on an analysis of student errors.

Swets and co-workers (1962) applied computer-based programmed instruction to examine the usefulness of overt response, immediate reinforcement, adaptive sequencing, and self-pacing in learning to identify multidimensional sounds. They found that these procedures "produced results that are comparable to those obtained previously with conventional training methods. Certain of the central features of automated instruction were found to hinder learning in the task studied." A description of one of their series of four experiments will illustrate method and results. The subject was presented with one of a

set of 32 stimuli over a pair of earphones, while seated in front of a typewriter connected to a computer. Each sound assumed one of two possible values along five dimensions: frequency, amplitude, interruption rate, duty cycle, and duration. Figure 4 shows the percent correct attained on an identification test after a one-hour training session with one of five procedures. In condi-

Insert Fig. 4 about here.

tion 1, at the subject's command the computer typed a five digit number that specified the stimulus and then played the sound; the subject did not emit an overt identification response but proceeded to give the next command. In condition 2, the subject initiated the presentation of the stimulus, then typed the five digit number he believed to identify the sound; the computer typed OK or WRONG and, if wrong, typed the correct answer. Condition 3 was like condition 2 except that, after an erroneous response the subject received the sound again and made a second attempt at correct identification. Condition 4 was like 3 except that, if the subject gave an incorrect answer, he heard the sound corresponding to it, as well as a repetition of the test sound before he made a second try. In condition 5, the number of trials under the condition 1 procedure was reduced by half so that an equal number of trials was received under all procedures. As a consequence of these procedures the probability of response confirmation varied from 0 under condition 1 to maximal under condition 4. The authors conclude:

"The data provide no support for the hypothesis that continual interrogation and overt responses are beneficial to learning; [they]

show that a fairly extensive feedback may be detrimental...and provide no support for the hypothesis that efficiency of learning varies directly with the probability of reinforcement."

The authors' conclusions must be generalized with caution. Negative findings in an evaluation of programmed instruction may be attributed to many sources. In second-language learning, these include faulty analysis of the terminal behaviors, false assumptions about the relevance of the experimental tasks to language behavior, and failure to revise the program iteratively, based on an analysis of student errors. Negative findings can be interpreted properly only when they are based on a program that has been shown, under other controlled conditions, to be effective.

Speech synthesizers such as pattern playback (Cooper, Liberman, and Borst, 1951), PAT (Lawrence, 1953) and OVE (Fant, 1958), make it possible to perform comparably well-controlled experiments in programmed learning with speech stimuli; regrettably, these devices have not found application in this area as yet.

Stimulus Generalization

Stimulus generalization, that is, the degree of invariance in responding under changes in stimulus conditions, has received extensive study in the laboratory with both non-speech and speech stimuli.³

An experiment by Cross and Lane (1962), employing stimuli varying along a single dimension (intensity of a 500-cps tone), is of particular interest because two incompatible responses were conditioned to two disparate stimuli

and then the probability of emission of each response to intermediate stimulus intensities was determined. This paradigm for discrimination learning and generalization testing bears some resemblance to the language-learning situation in which responses from the student's first- and second-language repertoires compete for emission. Figure 5 shows that response probability is maximal and latency minimal when the stimulus associated with reinforcement

 Insert Fig. 5 about here.

(S_1^D or S_2^D) is presented. When the probabilities of emission of the responses are equal, at a median stimulus loudness, the latency of responding is greatest. Curiously enough, the same findings were obtained when two compatible responses were employed (low-pitch vowel and high-pitch vowel); when response probabilities were equal, latency was maximal and there was no evidence of response "blending" i.e., intermediate pitch, under these conditions.

This experimental paradigm has recently been extended to multidimensional synthetic stimuli with essentially the same results (Cross, 1963). It turns out that, if a discriminative response is conditioned to a stimulus with the properties (a_1, b_1), the degree of generalization to another stimulus, differing along both dimensions, (a_j, b_j), is equal to the degree of generalization due to the change in a , ($a_j - a_1$), plus the degree of generalization due to the change in b , ($b_j - b_1$). It would be interesting to explore generalization in the two-dimensional stimulus array defined by the first- and second-formant frequencies of synthetic vowels. Departures from this simple additive rule of generalization might occur in the region of points which define vowels in

the subject's native language.

Over the past two decades, investigators at the Haskins Laboratories have greatly enhanced our understanding of the perception of speech by measuring the stimulus generalization of listeners responding to synthetic speech stimuli. Illustrative of a segment of this research is a study by Liberman et al. (1957) of the perception of the voiced stops /d/ and /t/. Spectrographic patterns were prepared and converted to sound by means of a device called Pattern Playback. The patterns were identical except for the relative onset times of their first and second formants: the first formant was cut back in 10-msec steps from 0 to 60 msec. They were presented in random order to subjects with instructions to label each stimulus as either /do/ or /to/. Cross and Lane (1962) replicated this study with six subjects, and also measured response latency; their findings are shown in Fig. 6. Since the

 Insert Fig. 6 about here.

/do/ and /to/ response-gradients are complementary, only the former is presented. The reader may wish to compare these data with those in Fig. 5; both experiments measure stimulus generalization after two incompatible responses have been conditioned to two discriminative stimuli. In studies of generalization among first-language speech stimuli, prior discrimination training has been arranged, of course, by the verbal community and is not controlled by the experimenter.

The average generalization gradient in Fig. 6 is more shallow than that for individual subjects. A very steep phoneme labelling gradient (in the ex-

treme case, a step function) means that the subject responds alike (generalizes) to all stimuli within a phoneme class and discriminates completely between stimuli drawn from the separate classes. In other words, discrimination training has effectively partitioned the stimulus continuum into two discrete classes of stimuli: do-stimuli and to-stimuli; that is, into two phoneme categories. Support for this conclusion comes from further experiments by Liberman and his co-workers with a number of phonemic contrasts. They find that a subject has great difficulty in distinguishing (in a psychophysical task) between those stimuli which he labels alike, i.e., those drawn from the same phoneme category, while he readily discriminates stimuli drawn from different phoneme categories. Particularly noteworthy is their finding that discriminability of the two phonemes is predictable from the labelling gradients. The analogous finding for a non-speech continuum such as pitch would be that the listener discriminates only as many pitches as he can name. This conclusion is in error by a factor of about 200 (the number of DL's for frequency at moderate intensity divided by the magic number 7).

Liberman (1957) argues that, in the case of speech stimuli, the physical continuum is broken into two discrete perceptual classes because two discrete responses mediate between the stimulus presentations and the recognition (labelling) or discrimination (same-different) responses. Thus, for the acoustic continuum in Fig. 6, stimuli differing in the delay of first-formant onset evoke covertly either the [do] response or the [to] response; these discretely different covert responses (voiced - voiceless)—or their neural correlates—then control the overt discriminations by the subject. This account of speech discrimination in terms of response differentiation, called the motor

theory of speech perception (Lisker, Cooper, and Liberman, 1962), has important implications for selecting a procedure by which these two components of formal repertoires shall be conditioned and coordinated (see below).

The experiments in stimulus generalization described above sampled the formal repertoire called dictation since non-vocal responses were under the control of an array of auditory stimuli. Since only two responses were employed in each of the studies, an extensive, differentiated repertoire of responses was not required. A number of experiments on generalization among speech stimuli that differ along several dimensions have capitalized on the extensive repertoire of vocal and written responses that is the largesse of first-language learning. With this technique, a confusion matrix may be obtained: the discrete stimuli are the column headings, the discrete responses the row headings, and the frequency of occurrence of R_i over the n presentations of S_i gives the entry in the i^{th} cell. Thus, a confusion matrix is a set of generalization gradients where, typically, the independent variable is on a nominal scale and the dependent variable is response probability (a ratio scale; vide Stevens, 1951).

A study by Sapon and Carroll (1958), who determined confusion matrices for selected words from English, French, Spanish, Portuguese, Italian, and Russian, illustrates this type of experiment. The procedure was similar to that later employed by Suppes, Crothers, and Weir (1962), described above. Japanese, American, and Latin-American listeners with virtually no second-language training were employed. The pattern of errors in phoneme matching was found to be nonrandom; significant differences among the three language groups of listeners were obtained. To cite one example, stop consonants

were always matched with other stop consonants by Latin-American listeners, while American and Japanese subjects often chose fricatives. In the light of these findings, the authors properly warn against interpreting confusion data obtained with one language group as generally characteristic of human sensory process or stimulus generalization.

As indicated earlier, these confusion matrices provide a valuable guide to the second-language programmer and an empirical check on the order of discrimination difficulty based on contrastive analysis. Reiff (1961) has suggested that this analysis would ideally contrast each student's idiolect with that of the target language. Short of this extreme, he mentions that the Linguistic Atlas of the United States may serve as a guide when the student's residence is known. For example, "a native of Eastern Pennsylvania will need a heavier preponderance of the discrimination [o'] [əw] - or - [éw] than between [o'] and [ów]." It does not seem implausible that programs to condition second-language echoic behavior, which are developed in different regions of the United States, may differ in their emphasis on certain discriminations.

Response differentiation

The differentiation of spoken or written responding, which is the second requirement of a formal repertoire, is **predicated on the inherent variability** of all behavior. In the laboratory, the technique of shaping or response-differentiation amounts to this: "We select one (or more) of the 'natural' **variations of a well-conditioned response and give it exclusive reinforcement;** the remaining variations are subjected to extinction. If we pick out, in advance, a variation that has been of fairly frequent occurrence, and if we ap-

ply this selective reinforcement rigorously, we can soon produce an increase in the frequency of the response that possesses the property or properties that we have chosen " (Keller & Schoenfeld, 1950).

Differential reinforcement of vocal responding has rarely been studied in the laboratory; this is particularly regrettable in view of its importance in the acquisition of second-language fluency and the relative ease with which the topography of this behavior may be measured (vide Lane and Shinkman, 1963). A study of shaping vocal duration, reported by Lane (1963b), illustrates the method and findings in this application. The subject is seated in an audiometric room in front of a microphone, light, and coin dispenser. He is told to say /u/ when the light flashes and that the only pay he will receive will be dispensed in front of him, contingent on appropriate behavior. Each of the first 20 responses receives reinforcement (1 cent) and its duration is noted. A subclass of this operant is then chosen for selective reinforcement. In the experiment whose findings are shown in Fig. 7 the mean (M) and average

 Insert Fig. 7 about here.

deviation (AD) of the set of 20 responses was computed and, in the second phase of the experiment, all responses whose duration exceeded $M_1 + AD_1$ were reinforced. The phase was continued until 10 successive reinforcements were received. Then, $M_2 + AD_2$ was computed for these 10 responses, and this value set as criterion for phase 3. In this fashion, the vocal duration of the subject was shaped (filled triangles) from approximately one-fifth of a second to nearly one-half second. After the fifth phase, the experimenter reversed

the direction of shaping, now reinforcing all responses with duration less than $M_5 - AD_5$. The experiment was terminated when the subject's vocal duration was reduced to one-tenth second. This subject, like most of the others whose behavior was similarly manipulated, was unable to report the property of his behavior that produced the money ("unable" in the sense that extensive prompting by a carefully-worded questionnaire, presented after the experiment, failed to yield an accurate report, if any at all).

Other experiments in this series varied the "size of step" that the subject had to take from one phase to the next; that is, the difference between the mean duration of ongoing and the new criterion, $(C - M)$, was varied. A significant generalization arose from the data for over 30 subjects and a wide variety of step sizes. It was not possible to predict, from the size of step alone, either the success of the shaping process, that is, whether the subject would reach criterion, or the amount of shaping necessary before criterion was reached. However, it was possible to predict the length and outcome of the shaping process when the step size $(C - M)$ was expressed in terms of the variability (σ) of ongoing behavior. The statistic $\frac{C - M}{\sigma}$, called the shaping index, was correlated 0.4 with the number of responses that were emitted before criterion was reached, and it was found to be large whenever the shaping process failed.

When the shaping index is large, either because the step size is large, the variability small, or both, the a priori probability of a reinforced response is low and the success of the shaping process is endangered. When the shaping index is small, the probability of a reinforced response is high, and the success of shaping is secured; however, this may require many small steps

to lead from the student's initial repertoire to the desired terminal repertoire. Therefore, the security of the shaping process and its efficiency are maximized when the probability of a reinforced response is set at some value, as yet unexplored, between zero and one, through a suitable choice of step size. It may not be too great an extrapolation of the initial findings of this research to suggest that the widely-discussed issue of step size in programming various subject matters should be viewed in this light: the optimal size of a step is best defined in terms of the behavior of the individual student just before taking that step, so that the probability of a correct response has some constant, intermediate value (cf. Geis, 1961).

The technique illustrated by these experiments in shaping response topography should provide a basis for measuring the interference produced by the first-language repertoire during shaping of second-language fluency. Just as the predictions of difficulty in discrimination-learning provided by a contrastive analysis may be verified by studies of stimulus generalization, so, too, these predictions for difficulty in response-differentiation may be verified by measurement of response topography during experiments in shaping.

An article by R. J. Sweet (1961) describes the application of contrastive analysis to the prediction of difficulty in response shaping, with particular relevance to sources of interference between English and French repertoires. "The control of pronunciation, then, involves the sequencing of phonemic elements and their combinations so as to minimize the possibility of confusing phonemes and their combinations in the new language with others already acquired in the native tongue. Sequencing of these elements depends upon the existence of varying degrees of difficulty in their pronunciation." The

author reports that pupils in his classroom research most often mispronounce those French words which are nearly approximated in English; e.g., papier and paper, borsse and brush, personne and person.

Several writers, among them Sweet (1962) and Hayes et al. (1962), have suggested that the shaping of new response topographies may be accelerated by employing those in the first-language repertoire. For example, the student may be led to a correct pronunciation of the German ich via the following steps: (1) "fish," (2) "fish" said through spread lips; (3) "fish" said through spread lips while opening the teeth slowly (after Hayes et al., 1962). The danger of such techniques is, of course, that they evoke the very responses in the first language which are sources of interference and must be weakened. Here the language programmer faces a choice, unguided as yet by research findings, which recurs in programming other repertoires as well (see textual behavior below). By calling into play first-language discriminations and patterns of responding, an approximation to the desired terminal behavior is more quickly attained, but it may then be more difficult to move precisely onto target.

Coordinating discrimination and differentiation

In any formal repertoire, there is a point-to-point correspondence between stimulus and response, whether these be spoken or written. The second-language programmer must undertake, after engineering discrimination and differentiation separately, to coordinate them. The mastery of International Morse Code illustrates the paradigm; there are 36 basic stimulus patterns each

connected with a differentiated response pattern. "When one learns to receive code, his problem is mainly discriminative, since the written or spoken responses have already been well-differentiated; in sending code, however, the problem is one of differentiation, since the discriminative work was done when the student learned to read his ABC's" (Keller & Schoenfeld, 1950). Note that conditioning a formal repertoire, such as receiving or sending code, establishes a minimal functional unit of behavior and permits the appropriate extension of this repertoire to new cases; this contrast between formal and thematic repertoires was cited earlier. In the case of sending Morse Code, there are probably 36 functional units; this "minimal textual repertoire" is readily extended—the telegrapher can, of course, send words from English text which he has never seen before.

A consideration of the development of echoic behavior will illustrate the problems of coordinating discrimination and differentiation in all four of the formal repertoires. We may restrict our attention to strategies for conditioning a minimal echoic repertoire which is later extended to new stimuli as the occasion demands. This minimal repertoire is vital to second-language learning for, once it is conditioned, the programmer can evoke new responses as an assemblage of small echoic units never before arranged in this order; this behavior may then be reinforced.

When programming echoic behavior in a second-language, certain auxiliary behaviors by the student, which are normally taken for granted in other kinds of programming, are not available. In the more common "subject matter" programs, confirmation is contingent upon not only a correct response but also upon a correct discrimination by the student; namely, that his written re-

sponse "is the same as" the printed answer. This requirement of a discrimination is rarely mentioned and remains unobtrusive until the student erroneously considers equivalent such response-confirmation pairs as, "can not" and "does not" (Holland, 1959). Under these circumstances, the confirmation stimulus does not serve to reinforce the correct response.

The requirement of a discrimination for self-reinforcement obtrudes from the start, however, in programming second-language acquisition, for these discriminations are not part of the initial repertoire of the student. In conditioning echoic behavior, the programmer's task is further confounded by the fact that the confirmation stimulus and the subject's response are both fleeting and presented via different modalities; an aural stimulus in the first instance, sidetone and bone conduction in the second. Furthermore, psychophysical differences in the perception of external vs. "autophonic" sounds probably impede the development of accurate echoic behavior (vide Lane, 1962b).

How then shall the student's echoic accuracy be judged? Assuming that each student shall not have access to a private tutor or to an automatic speech recognizer, there seems to be little choice but to condition the prerequisite discriminations in the first place, and arrange that they are transferred to later conditions of self-shaping echoic behavior. This widespread practice is based more on a pious hope than a solution. For example, Pimsleur (1963) writes:

"The popular notion is that the student who is able to compare his pronunciation with that of a native speaker will acquire native-like speech. Many teachers have already realized that this hope is

far too optimistic. It is evident that students are rather poor judges of their own pronunciation. They are prone to think their pronunciation is "good enough" when actually it is not acceptable. They are unable to note which features are relevant and which are not."

There is little experimental evidence to suggest that the relevant discrimination, once conditioned, are effective in self-shaping echoic behavior. Furthermore, few of the programs that adopt this strategy provide for maintenance of the initial discriminations during self-shaping in later stages of the program.

A series of 11 pilot studies and three major experiments by Pimsleur, Mace, and Keislar (1961) are among the few that examine the effects of discrimination training on response differentiation. The central problem of their research was "to determine the value of preliminary discrimination training in increasing the effectiveness of language laboratory practices on the pronunciation of French sounds." They found: "(1) Training in discriminating the French nasal vowel phonemes /*õ*/, /*ã*/, /*ẽ*/, rendered subsequent language laboratory practice more effective in producing good pronunciation of these phonemes. (2) Training in discriminating between the diphthongized and the undiphthongized final /*o*/ did not render language laboratory practice measurably more effective in producing good pronunciation of this French phoneme." Note that the first discrimination is among phonemes in the target language, while the second is between a French phoneme and a troublesome English approximate. Pimsleur (1963) attributes the different outcomes to the greater degree of response differentiation that existed in the former case for native

speakers of English.

A close relation between speech discrimination and differentiation is again implied by the outcome of a programming experiment in speech correction. With a carefully designed and executed program that should serve as a model for others, Holland and Matthews (1963) taught [s] discriminations to children who misarticulated [s]. The discriminative response was button-press; correct responses simply led to the next item; incorrect responses caused the tape recorder to rewind and repeat the last item. Phase 1 of their program involved discrimination of [s] in isolation from other isolated speech sounds. The latter were sequenced according to difficulty and faded in toward equal intensity with the S^D. In Phase 2, the subject was required to discriminate the sound in words. Sequencing of difficulty and fading were again employed. For example, words with initial [s], terminal [s], and medial [s] were presented in that order. In Phase 3, the child was asked to identify the position of [s] within a word. Phase 4 involved discrimination of correctly articulated from misarticulated [s] sound within words; omission, substitution, and distortion comprised the misarticulations.

The group of children who worked through this program showed significantly improved [s] discrimination, sibilant discrimination, and, of particular interest here, improved [s] articulation on a series of tests.

There are several considerations, however, that strongly question the value of discrimination training before response differentiation. The first of these is the outcome of an experiment by Lane and Schneider (1963) that measured the accuracy of self-shaping echoic responding to Thai tonemes under several procedures. These authors report that discrimination training, in which

the target toneme was contrasted with segments of the same form but different durations and pitch slopes, did not lead to a marked improvement in echoic accuracy. The further introduction of delayed auditory feedback of the subject's response, facilitating comparison of the model and his attempt at imitation, also failed to produce an appreciable improvement in the correspondence of stimulus and response pitch or duration. Nevertheless, the subjects had previously shown perfect discrimination between tonemes that were more similar than the model and their attempts at imitating it. This high level of discrimination was reached rapidly during training and probably reflected transfer of first-language discriminations, as discussed earlier in this article in the section on Discrimination. The findings reported in that section, taken together with these data on echoic behavior, indicate that inadequate response-differentiation can persist despite a very fine-grained discriminative repertoire.

Research at the Haskins Laboratories on the generalization of synthetic speech stimuli, discussed earlier, casts further doubt on the necessity of discrimination preceding differentiation in the conditioning of formal repertoires. Indeed, inference from these findings suggests that discrimination training ideally should follow response differentiation. In a review of research on speech perception, Liberman (1957) writes: "All of this strongly suggests...that speech is perceived by reference to articulation—that is, that the articulatory movements and their sensory effects mediate between the acoustic stimulus and the event we call perception." It follows from this motor theory of speech perception that two relatively similar external stimuli could be made more discriminable "if we could attach to those stimuli two very

different mediating responses and hence gain the added distinctiveness of their very different proprioceptive returns."

As an example of this suggestion we may consider the implications of the studies by Liberman et al. (1961) on the discrimination of /p/ and /b/ in intervocalic position. An experiment similar to the /do/ - /to/ study described earlier was performed with the synthetic speech stimuli, *rabid-rapid*. These stimuli varied only in the duration of the silent interval between the two syllables in each word, which ranged from 20 to 130 msec. Labelling responses by native Americans "partition" the stimulus continuum into two perceptual classes: call them rabid stimuli and rapid stimuli. This outcome is attributed to covert mediating responses which provide discrete cues (e.g., voiced-voiceless) for the labelling judgments. A modest extension of Liberman's hypothesis suggests that a foreign student who has not learned the differentiated response patterns /p/ - /b/ in English should find this auditory discrimination much more difficult than do native Americans. Moreover, if the new responses /p/ and /b/ are differentiated out of the student's current vocal repertoire, this should enhance the discrimination of /p/ and /b/ when presented to him as auditory patterns, for now the student's responding may be controlled by intervening articulations that provide very different proprioceptive cues.

Further support for the suggestion that vocalization may facilitate the development of speech sound discrimination comes from some observations reported by Holland and Matthews in their study of [s]-discrimination, discussed above. The authors actually compared several programs for conditioning this discriminative behavior; they note that the one described earlier

produced the most spontaneous vocalizing by students working through the program as well as the greatest improvement on the [s]-discrimination tests. They suggest that "these vocalizations might force careful observation of auditory cues, or they might provide supplemental kinesthetic stimuli which could be useful in close discriminations."

It is unfortunate that the pedagogical implications of the motor theory of speech perception have not been put to empirical test. The outcome might have importance not only for second-language learning but also for an evaluation of the theory itself, which presently lacks a firm empirical basis (vide Lane, 1963a).

1. Echoic Behavior

Among the four formal repertoires delineated earlier, echoic behavior, involving a spoken response to a spoken stimulus, receives the major emphasis in modern-language pedagogy. "Today we are pretty much in agreement that the grammar-and-translation method, patterned largely on the traditional study of Latin, is obsolete at the least" (Mildenberger, 1962). Instead, the major emphasis is on aural presentation of stimuli and oral response—the audio-lingual method. The objectives of the method include thematic repertoires to be discussed later, but its modus operandi involves the purely formal components of an echoic repertoire. "The justification for this emphasis is found in the observation that a language is first of all a system for social communication; writing is a secondary derivative system for the recording of spoken language..." (Carroll, 1962a).

The audio-lingual method may be characterized by Brooks' (1959) well-

known dicta: "1. The learner should hear only authentic foreign speech. 2. He should hear much more than he speaks. 3. He should speak only on the basis of what he has heard. 4. He should read only what he has spoken. 5. He should write only what he has read. 6. He should analyze, if he does so at all, only what he has heard, spoken, read, written, and learned."

Despite the widespread use of the audio-lingual method and the related proliferation of language laboratories (Hayes, 1963), considerable uncertainty remains about the value of this approach to second-language learning. Questions such as these, posed by K. A. Mueller (1962), are largely unanswered by available research findings: Does withholding texts, and the absence of reading and writing, in the early stages of language learning by this method affect performance in pronunciation, oral comprehension, reading, writing and syntax at various stages in the course? Does the use of the language laboratory contribute to performance on audio-lingual skills and on others, too? What types of language laboratory facilities are most valuable (vide Morton, 1961b) in producing second-language fluency?

Scherer and Wertheimer (1962) summarize the dispute over the audio-lingual method:

"Inherent in the assumptions underlying an audio-lingual approach to foreign language learning is that students trained only audio-lingually at the beginning will subsequently reach, and perhaps surpass, in reading and writing ability, students who have been trained in the more conventional manner, while exhibiting a far higher level of achievement in auditory comprehension and speaking; advocates of the traditional multiple approach, with its heavy emphasis on grammar and read-

ing, are of a contrary opinion. Little evidence based on controlled experimentation exists."

These authors have performed a rigorous and comprehensive experiment to contrast the effects of the audio-lingual method, which employs primarily echoic behavior, to those of "traditional" methods, which rely more heavily on textual behavior. Noteworthy among their criteria were several novel tests of "habituated direct association" (roughly, "the ability to think in German"), which are described by Wertheimer and Scherer (1962). To précis their findings: audio-lingual students were superior in listening and speaking ability, comparable in reading and writing, and inferior in translating when compared to a control group taught by traditional methods. The former were also superior in responding to spoken German rapidly, appropriately, and in a fashion comparable with that of German bilinguals. (Other comparisons of traditional and audio-lingual methods are cited by Carroll, 1962a.)

The third tenet of the audio-lingual method—the student should speak only on the basis of what he has heard—may be reflected in the decision of many language programmers to condition discrimination before response differentiation. The decision cannot be based presently on laboratory research, for the findings, discussed in the earlier section on Coordination, are equivocal. Sapon (1961a) incorporates the audio-lingual principle in his program in elementary Spanish: "In evaluating any program, one must remember that an individual can say only what he can hear, and one of the great strengths of this program is its training in auditory discrimination." Morton's (1960) audio-lingual program for Spanish also has this approach. His first "hypothesis" and procedure in the program is called "Phonematization":

"Essentially, this would hold that the incipient language learner learns first to 'hear' and discriminate between all 'significant' classes of sounds in the new language before a conscious effort is made to reproduce them; or, indeed before any real capability exists for reproducing them. It is an aural conditioning necessary to set up the 'phonemic areas' which will work finally in his acoustico-neurological system like a series of narrow-band filters, passing appropriate sounds, rejecting others automatically..."

When discrimination learning is accomplished, most second-language programs move on to echoic behavior, in the hope that the prerequisite response-differentiation will literally take care of itself (through some mechanism of self-shaping). For example, in Morton's audio-lingual Spanish program the procedure of "Phonematization" is followed by one called "Sound Reproduction":

"...automatic reproduction of all speech sounds as individual acoustic segments is achieved after phonematization of these sounds...By means of the 230 individual sound reproduction drills, the student was expected to learn to render correctly, both in isolation and in combination, all of the sounds and sound groups (syllables and utterances) phonematisized earlier."

Similarly, programs in elementary French by Valdman (1962) and by T. Mueller (1962) begin with discrimination training, followed by response differentiation through echoic behavior. In Mueller's program, each vowel and consonant defines a separate unit in discrimination training. Each unit "teaches the student to discriminate against the S^A sounds and, after that, against the various

intonational features such as: pitch, stress, rhythm, and syllabification."

Self-shaping of echoic behavior follows this discrimination training.

Lermontoff (1961) advocates the early introduction of textual material in programming echoic behavior. He argues that "some subjects find it easier to discriminate two sounds after they have read words in which they occur." Consequently, this author's program in elementary French has the following features:

"We first present a new word aurally; and next we have the subject listen to it again and repeat it; then we have him listen to it and repeat it while he is at the same time presented with its written form; and finally we have him read the word aloud and then let him hear the correct pronunciation on the tape."

A French program by Moraud (1961) also integrates echoic and textual behavior. The conditioning procedure has six steps.

1. [The students] hear a sentence...

...and read the text presented...unless the text has been faded out completely. 2. They repeat the sentence into the microphone or give answer into microphone 3. They write answer on answer sheet unless instructions call exclusively for an oral response. 4. They listen to correct oral response and read correct written response. 5. They make appropriate connections. 6. They move to the next frame."

Rocklyn (1961) and his collaborators have prepared programs in several languages to condition echoic behaviors concurrent with a limited thematic repertoire. Their effort is of particular interest because they have applied

linguistic and behavioral analysis to the choice of terminal behaviors as well as to the design of a program. Usually, the second-language programmer must accept the target language as given. However, programming should be viewed as one technique among many in educational technology and, whenever it is possible to do so, the entire learning task should come within the purview of the behavioral engineer. A report by Rocklyn and Moren (1962) describes programming technique of this group as applied to the acquisition of a limited tactical language in Russian. The language was designed to include 16 commands (e.g., "hands up"), 10 question frames ("how many"), 100 inserts ("tanks, guns"), and 100 answers (yes, no, numbers, etc.). The conditioning paradigm included a recorded Russian or English stimulus, a student response in either Russian or English, depending on the stage in the program, and a recorded correct utterance as confirmation.

Although the acquisition of echoic behavior overlaps with that of thematic repertoires in most second-language programs, in a few the two repertoires are strictly sequenced. For example, Morton (1960) writes, concerning his Spanish program, "...automatic reproduction of all speech sounds....is achieved...before the consistent use of the sounds in meaningful utterances," and Valdman (1962) reports on his program in elementary French: "We begin with a sound discrimination and production sequence followed by grammatical transformations dealing with features comprised exclusively of sounds that the student has been taught to discriminate and differentiate."

To summarize: Echoic behavior receives the greatest emphasis among the formal repertoires, partly because of the widespread acceptance of an audio-lingual approach to language pedagogy in general. Most often, the development

of echoic behavior skips from discrimination training to imitation with the intervening steps—response differentiation and coordination—largely omitted. In so doing, the programmer capitalizes initially on the students differentiated first-language repertoire, leaving to the self-shaping process the resolution of phonetic differences between the two languages. Echoic behavior is taught before thematic repertoires in some second-language programs, and concurrent with them in others.

Several specialized devices have been designed to implement programmed acquisition of echoic behavior. Gilpin (1961) has described a system composed of two tape recorders (master and student) and relay circuitry which is operated by the student with a control panel. Fifteen buttons permit control of the two recorders in a variety of ways that provide an extremely flexible arrangement of stimulus and response recording and presentation. Carroll (1961) has designed and constructed an audio-visual device, shown in Fig. 8. The device

 Insert Fig. 8 about here.

has three operational modes. In the "familiarization mode," the student is first presented with instructions, new material and the like; then he is given some task to perform, such as answering a question; then he finds out whether the response he made was correct or not. There is provision for vocal, written, and multiple-choice (button-press) responding. When the student meets a criterion performance in this mode he proceeds to the "learning mode" which omits instructions and provides only questions and confirmation. After criterion is reached in this mode, the student goes on to the "test mode" where only

questions are presented.

Lane (1963c) has designed a device to permit programming and selective reinforcement of echoic behavior restricted to the prosodic features of speech. This device extracts the intonation, tempo, and relative amplitude parameters of a recorded stimulus pattern and of the subject's echoic response, compares one or more parameters for the stimulus and response, and presents either an analog or discrete representation of the parameter mismatch to the subject. Depending on the magnitude of the error voltage and a programmable criterion, the subject is automatically presented either with a repeat of the last pattern or with the next pattern in the program.

Porter has surveyed the use of audio-visual aids (1958) and of other instructional devices (1960) in foreign language teaching with emphasis on the implementation of programmed instruction. Carroll (1962b) provides a list of teaching machines designed for second-language programs.

2. Transcription

When both the stimulus and response are written, all the properties of echoic behavior apply in the conditioning and maintenance of transcription. If the second-language orthography is not common to English, the development of a purely formal repertoire will require both discrimination training and response differentiation. Such problems as the minimal unit, extension to new stimuli, the "grain" of the formal repertoire, mimicry, and automatic self-reinforcement, which arose in our discussion of echoic behavior, are pertinent to transcription and their solutions are analogous. For example,

because the stimulus and the response-product are commensurate in transcription, as they are in echoic behavior, a possible programming strategy is to condition orthographic discriminations and responses initially, and then trust that the student will discriminate the accuracy of his transcriptions, automatically shaping his own behavior to minimize the mismatch between his response and the model pattern. If the second-language orthography is largely the same as in English, the required stimulus discriminations and response differentiation have already been acquired. This is the case for most second language programs. Noteworthy exceptions are the programs that condition transcription in Russian (Saltzman, 1962) and Hebrew (Hammond, 1962).

3. Textual Behavior

Textual behavior is vocal responding under the control of a non-auditory verbal stimulus. The emphasis here is on the formal properties of such behaviors as reading aloud from a second-language text in that language. Formal and thematic repertoires are combined when a student reads a second-language text while translating into his native language. Many of the problems in programming the acquisition of textual behavior in a second-language are similar to those associated with echoic behavior and transcription discussed earlier. Discriminations among the second-language orthography must be conditioned and vocal responses must, of course, be shaped out of the student's initial repertoire. If the target language employs fundamentally the same orthography as in English, these discriminations require only minimal attention from the programmer. Furthermore, if textual behavior is programmed

following echoic behavior, then the requisite response differentiation is available without further conditioning. It remains only to set up correspondences between discriminative stimuli, usually printed text, and differentiated vocal responses.

This strategy in programming corresponds to the fourth tenet of the audio-lingual method—the student should read only what he has spoken—and it has a counterpart in the classroom practice of teaching audio-lingual skills before reading. As Scherer and Wertheimer (1962), K. A. Mueller (1962), and others have pointed out, there is little research evidence in support of this practice. Indeed, when audio-lingual methods have been compared with sight methods in producing speed and retention in reading, all four possible outcomes have been reported: (1) "Oral-aural and reading proficiency constitute separate, independent skills which do not develop one from the other" (cited in Carroll, 1962a); (2) "aural learning facilitates visual relearning, this facilitation is greater than the facilitation of aural relearning by visual learning," (Pimsleur & Bonkowski, 1961); (3) "subjects who learned visually and relearned aurally achieved superior performances in comparison with performances of subjects who learned aurally and relearned visually," (Asher, 1962); (4) "students taught by the audio-lingual method...were almost on the same level with students taught by the traditional method in new-type reading and writing tests" (Scherer & Wertheimer, 1962).

The greatest difficulty in programming textual behavior arises from a difference between this repertoire and transcription or echoic behavior. The latter repertoires have in common that the response-product and the stimulus occur in the same modality, visual in the first case, auditory in the second.

Hence, it is possible to discriminate an improper response by direct comparison with the stimulus. On the other hand, dictation and textual behavior have in common that the stimulus and the response product are dissimilar—they impinge on different modalities. Thus, penmanship can be learned by transcription but not by dictation alone, and pronunciation can be learned from echoic behavior but not solely from textual behavior. Although there is a point-to-point correspondence between stimulus and response in textual behavior, there is no formal similarity between the two which, coupled with speech discriminations, could maintain response differentiation. Skinner (1957) draws the contrast clearly:

"In echoic behavior, the correspondence upon which reinforcement is based may serve as an automatic conditioned reinforcer. The speaker who is also an accomplished listener 'knows when he has correctly echoed a response' and is reinforced thereby. Such reinforcement brings the form of the response closer and closer to the form of the stimulus, the limit being the most precise correspondence possible either with respect to the vocal capacity of the speaker or his capacity to judge similarity...The automatic reinforcement of reading an 'interesting' text, however, has merely the effect of increasing the probability of occurrence of such behavior; it does not differentially reinforce correct forms at the phonetic level."

Although the speaker who is reading text is not able to assess the accuracy of his pronunciation by direct comparison to a model, he is, nonetheless, able to listen to his own speech and this probably has some editing

effect on gross mispronunciation. It would be interesting to determine whether the maintenance of response differentiation in textual behavior would be impaired by removing the auditory feedback to the speaker, as through masking of sidetone. The difficulty of maintaining differentiation in textual behavior need not be a problem for the second-language programmer who desires only to condition "reading a foreign text for understanding." The widespread practice of conditioning students to respond aloud to text in a second language with accurate pronunciation, or with the English approximations that the text evokes through generalization with English orthography, seems most inefficient when only translation of text is the goal. Most efficient would be a procedure which brought English vocal responses directly under intraverbal control (q.v.) by the text; no responses in the second language need intervene covertly. A formal repertoire is, of course, sacrificed by this procedure.

Although the acquisition of discriminative control in textual behavior is expedited when the student's first and second languages have largely the same orthography, the maintenance of the differentiative component of textual behavior is hampered by the same tokens. Sapon (1952) advocates the elimination of orthography in the early stages of language learning for this reason. He cites as an example the difficulty of teaching an English student learning Spanish to forsake a lateral continuant in response to the textual stimulus "r" (the product of some years of conditioning) and to emit instead an alveolar vibrant. On the other hand, Lermontoff (1961) was quoted earlier as favoring the introduction of textual stimuli during the conditioning of echoic behavior. The question, unresolved by research, is: does a common orthography

in the two languages facilitate the acquisition of textual behavior, by rapidly yielding an approximation to the desired response, more than it impedes this acquisition, by interfering with accurate pronunciation through the evocation of competing responses in the native language.

Sweet (1962) disowns the problem in his Latin program, where response differentiation is of little concern. Programs for languages whose orthography is not common to English, as is the case with the Cyrillic, Hebrew, and Arabic alphabets and Chinese, Japanese, and Korean characters, need not concern themselves with this problem but face another: discriminations among the orthographic symbols must be conditioned. (Saltzman, 1962; Hammond, 1962; Carroll, 1961).

4. Dictation

Dictation in a second-language is a formal repertoire that requires the conditioning of speech discriminations as in echoic behavior and, less often, response differentiation as in some cases of transcriptive behavior. The unit of the minimal repertoire may be orthographic, stenographic, phonetic, etc. The lack of formal similarity between stimulus and response provides several of the problems met in our discussion of textual behavior.

Among the programs that condition dictation, those by Morton (1960) and Harms (1962) bear special mention. Morton conditions dictation as a part of discrimination training; the responses are arbitrary orthographic symbols, called *grammalogs*; they are transitional behaviors, eventually replaced by differentiated vocal responding during the conditioning of echoic

behavior. Harms has developed a program to take a student from no knowledge of phonetic notation to a level of competence whereby he is able to give the phonetic notation at 95% accuracy for words selected at random from the Kenyon-Knott dictionary. The program proceeds from dictation of simple sounds through diphthongs and affricates, voiced-voiceless oppositions, stop-nasal oppositions, vowels in CVC frames, consonants with common places of articulation, semi-vowels, voiced and voiceless consonants, stress, and randomly selected units. The procedure is as follows: the student hears the word on tape, consults a corresponding card for prompting, writes the phonetic notation for the entire word, checks his response against confirmation on the back of the card, and repeats the sequence. Harms presents evaluation data for native and foreign-born students who worked through the program.

Thematic repertoires

Purely formal repertoires are of little functional value in their own right either for the speaker (or writer) or his audience.⁴ By virtue of the point-to-point correspondence between stimulus and response that characterizes these repertoires, the behavior is highly predictable and, in that sense, conveys little information. It is only when formal and thematic sources of strength combine in determining a response that we say it is meaningful. Because it has the proper form the response is, first of all, "intelligible"; that is, it has a form appropriate to the discriminative training of the audience. Because the response is also under the control of

environmental conditions, or operations on the speaker, or his own prior verbal behavior, or that of other people, or several of these acting concurrently, it "tells the audience about" these controlling variables and is likely to be reinforced.

In formal repertoires there is a point-to-point correspondence between stimulus and response—within one dimensional system for echoic and transcriptive behaviors, across two such systems for textual behavior and for dictation. In thematic repertoires, the response shows no such correspondence to the controlling variables.⁵ The term thematic repertoires will be used broadly here to encompass several functional classes of verbal behavior including intraverbal responses, which are under the control of prior verbal behavior, tacts, which are under the control of nonverbal environmental stimuli, and mands, which are under the control of characteristic reinforcing consequences (vide Skinner, 1957).

Although it is useful to distinguish the controlling relations involved in intraverbals, tacts, and mands, it is important to note that most verbal responses are multiply-determined by several sets of variables acting in concert. For example, the response "fire" may be a tact under the control of a conflagration in one's apartment, or it may be a mand for a characteristic reinforcer ("somebody sound the alarm!"), but it is probably both. It should also be noted that the same "word" may appear in one or more functional classes (in addition to a mand or a tact, "fire" could be an intraverbal to "Ready, aim,...").

5. Intraverbal Responses

Intraverbal relations within one language are variously called syntactic frames or sequences, grammatical structures, redundant strings, contextual constraints, etc. Skinner (1957) treats these under the special heading "autoclitics" in a functional analysis of verbal behavior. The most noteworthy feature of these chains is, of course, their patterning. Grammatical models, among them Markovian, immediate constituent and transformational, describe this patterning by different analytical techniques. Although these models are of interest to the second-language programmer, especially insofar as they describe the terminal repertoire that the program is designed to produce, a functional analysis of intraverbal chains is of pre-eminent importance since the programmer's task is primarily manipulative.

Unfortunately, there is a great dearth of laboratory research on the analysis and control of intraverbal chaining. Classic studies by Miller and Selfridge (1953) and Miller (1958) describe the potent effects of patterning in recall; recent psycholinguistic studies by Berko (1958), Brown and Berko (1960) and Brown and Fraser (1962) describe this patterning in the language of children; Miller (1962) reports several experiments on the effects of patterning on speech perception; techniques for conditioning and maintaining intraverbal chains, however, have not been examined experimentally. Saporta (1963) has undertaken an evaluation of three grammatical models in the teaching of foreign languages which should prove illuminating in this respect. Sapon's "reflections on models of linguistic structure and language programming" (1961) reflect, as well, the synthesis of analytical and functional approaches to patterning adopted by many language programmers:

"What has generally emerged from my own experience in programming Spanish is the recognition that the straight 'economical' line of language description [in linguistics] is neither the shortest nor the most direct route to student performance in the language being taught. What does seem to offer much promise is a first approach to the language via a series of transformations and expansions of basic sentence material. This material can then be sequenced to exploit usable existing habits and to avoid conflicting ones...a kind of contrastive psycholinguistic grammar...and finally fixed through the use of pattern and substitution drill."

In view of the widespread use of drill to condition intraverbal patterns, it is unfortunate that this technique has not been put to careful laboratory test. Under controlled conditions, it would not be surprising to find sheer repetition minimally effective. This technique should be contrasted with that of programming, in which gradual progression and the development of abstraction through multiple examples replaces repetition. Skinner (1960) has contended that no one learns by repeating: "he may learn so little that he needs to repeat and will learn more upon successive occasions, but the repetition itself is not involved." "Stamping in" or "fixing" through repetition alone has not been found to be an effective technique for the control of behavior in the laboratory.

Intraverbal relations between first and second languages comprise translation for the beginning student.

"Faced with a passage in the new language, the translator emits appropriate intraverbal responses. If these fall into something like

a familiar pattern, he may then react in any or all of the ways appropriate to a listener...Eventually the translator improves upon this crude procedure by developing more efficient intraverbal operants, mainly of larger patterns, and by acquiring normal listening or reading behavior under the control of the new language without the aid of translations" (Skinner, 1957).

Several language programmers agree that, whereas the evocation of formal repertoires in English interferes with the development of these repertoires in the second language, the evocation of intraverbal relations between the languages hinders not only the maintenance of formal repertoires in the second language but also the development of intraverbal chains in that language. Mueller (1961) writes: "Developing automatic control over the sound patterns of the language is retarded when the student's attention is directed toward meaningful utterances." A similar statement by Morton (1960) has been quoted earlier. (Also see T. Mueller, 1962.)

On the other hand, some programmers have argued for and implemented the early introduction of thematic repertoires such as translation (intraverbals) and question-and-answer (mands), on the grounds that motivation is sustained or enhanced by this procedure. Presumably, the reinforcing consequences of acquiring an intraverbal repertoire such as translation derive initially from the fact that the student may react to his own speech in his native language as a listener.

It remains to be shown by controlled experiments whether the introduction of thematic repertoires does in fact impair the formal repertoires established earlier; if interference is demonstrated, we need to know the

sources and their relative importance. The greatest difficulty in this kind of research is accurate specification of the dependent variable—the accuracy of the formal repertoire. Subjective techniques of evaluation are undesirable but acoustic specification of the stimulus and the response is at present unwieldy. It may not be premature, however, to undertake to assess the effects of introducing thematic repertoires on the prosodic features of formal repertoires, measured electronically.

Any of the four formal repertoires may serve as the vehicle for the acquisition of thematic repertoires such as intraverbals. Some programs employ a written stimulus and written response. These include a program in Hebrew by Hammond (1962), one in German by Ferster and Sapon (1960), and a computer-based program to condition Russian vocabulary by Licklider (1962). Valdman (1962), Moraud (1961), Harris (1961), and others condition intraverbal responses by means of dialogues supplemented by English translation. A Latin program by W. E. Sweet (1961) employs a less common procedure: a written stimulus is connected intraverbally to a spoken response. Most of the available second-language programs, however, condition spoken second-language stimuli to spoken English responses. Morton describes the procedure employed in his Spanish program as follows: "Practice rather than theory guided us in assuming that once the structural clues of Spanish had been mastered...the student, working with the minimal meanings afforded him by these in each new utterance, could be expected to 'fill in the gaps' with the help of 'broad hints' in English (i.e., nonliteral translations)."

Although most language programmers and teachers would agree that literal translation is not the solution to the problem of how to introduce an intra-

verbal repertoire most effectively, the available alternatives also seem unsatisfactory; these include non-literal interlinear translations, alternate pages of English and second-language texts, tape-recorded second-language phrases with their English "equivalents" in the text, paired associate learning with flash cards or lists, and so forth. Morton's "broad hint" approach, cited above, shows greater promise; it is related to a variably-blurred prompting technique, developed by Isreal (1960) to study vocabulary learning. The auditory analog of Israel's technique would be worth exploring: English prompts might be recorded on a tape track parallel to one containing foreign-language narrative or questions. The prompts could be masked by a noise whose level was under the student's control, or the prompts could be faded over the course of the recorded program by the experimenter, or both.

The crucial difficulty with all these techniques, however, is the attempt to transfer a response from one functional class to another. Initially, the student says "apple," for example, as a simple and well-conditioned textual or echoic response to the English "apple." It is desired, however, that this response come under the control (be evoked by) a formally dissimilar stimulus, e.g., the spoken or written "manzana." That association has no status in the student's native language. It is not surprising, therefore, that as long as the translation is available the English stimuli abrogate control of the English responses. Without the translation, however, there is no reason initially for the response "apple" to occur in the presence of "manzana."

One solution to the apparent paradox is to evoke the English responses by means of collateral stimuli that do not exercise extensive control over these responses, and then to fade these collateral sources of strength as the

intraverbal linkages between the two languages become established. This formula seems to characterize the ingenious technique developed by Schaefer (1961) to condition English intraverbal responses to German text. Among the collateral sources of strength for a response to English text, there are, in addition to thematic prompts, intraverbal prompts provided by the syntactic structure and by common phrases, as suggested earlier. These collateral sources of strength are nicely illustrated by the procedure for generating order-approximations to English with successive subjects, employed by Miller & Selfridge (1953). The redundancy of the text is one way of expressing the degree of control exerted over responding by collateral sources of strength. Consequently, when Schaefer replaces high-redundancy words in the English text with their German counterparts, as in "But why will you say that ich am mad?," the correct intraverbal is emitted and it is emitted in the presence of the German textual stimuli. It remains to fade these collateral sources of strength by replacing both more words and words of lower redundancy. For example, the above excerpt is soon followed by "Ich heard all things..." where "I" is less determined by context. In six pages the text works up to "Es war laut, aber es grew Lauter, ich say, Lauter every moment!"

The auditory analog of this procedure for conditioning intraverbal responses would be worth trying and might prove superior on at least two counts. First, Schaefer's procedure evokes English pronunciations of the German words, because of the common orthography employed and because German words are interspersed at varying intervals among the English words in the text. If echoic rather than textual behavior were employed, there would be less tendency for English pronunciation of the German words to occur and,

because the stimulus and response-product in echoic behavior are formally similar, it would be possible to discriminate the improper pronunciations. This is not possible in textual behavior, as emphasized earlier. The second potential advantage of an auditory adaptation of Schaefer's technique is the greater number of collateral sources of strength that may be manipulated. In addition to the intraverbal and thematic cues available in text, there is a multiplicity of acoustic cues, including stress, intonation and tempo, which may be used as prompts and vanished as necessary.

6. Tacts

The tact is a class of verbal operant characterized by a three-term relation: nonverbal discriminative stimulus, verbal response, generalized reinforcement.

"The three-term contingency in this type of operant is exemplified when, in the presence of a doll a child achieves some sort of generalized reinforcement by saying doll; or when a teleost fish, or picture thereof, is the occasion upon which a student of zoology is reinforced when he says teleost fish" (Skinner, 1957). Tacting comprises a large part of the thematic repertoire in first-language learning and a minute part in second-language learning; in the latter case, intraverbal responses "carry the thematic load" (vide Brethower, 1961). Instead of conditioning the response "manzana" with some generalized reinforcer in the presence of an apple or a picture thereof, the response is conditioned as an intraverbal to textual or auditory stimuli, such as: "Say 'apple' in Spanish," or "Eva comio una manzana envenenada." Once again, the means by which a thematic repertoire is taught

in second-language learning is to establish one sort of functional control and hope for transfer to another. Thus, when the student eventually confronts an apple and it is appropriate to say "manzana," we do not expect him to be speechless in the face of this novel situation; we expect manzana-the-intraverbal to become manzana-the-tact. The transfer of functional control presumably occurs through mediated generalization; that is, at the first confrontation the tact "apple" is emitted covertly, and "manzana" is uttered aloud as an intraverbal to "apple." If this leads to some sort of reinforcement on a number of occasions, "manzana" will acquire the functional properties of a tact.

The mediating responses in the student's native language, required by this approach to the conditioning of tacts (i.e., transfer from intraverbal control), probably impair the acquisition and maintenance of second-language fluency. Carroll (1962a) concludes in a survey of research that "foreign words are best learned (and probably better retained) when presented in association with the objects, actions, qualities, and conditions which are their referents." Several programs employ pictorial material to this end (e.g., Marty, 1962); much greater versatility is provided by educational television (Barcus et al., 1961) and specialized audio-visual devices (Carroll, 1961; Porter, 1958).

The direct approach to conditioning second-language tacts, however, also faces several problems. First, it is difficult to maintain these tacts outside the classroom, where the student responds to his environment in his native language. Second, the nonverbal controlling stimulus may evoke English responses that will interfere with both the formal and thematic proper-

ies of the second-language tact.

The nature of thematic interference between first- and second-language tacts is well-revealed in an experiment by Lenneberg (1957) on learning color names in six different experimental "languages." The languages consisted of four color words and the subject's task was to discover the "meaning" of the words by observing their use by the experimenter. The nonsense words of the first language were distributed over the Munsell color series exactly as the English words: brown, green, blue, and pink. The other five languages had words whose frequency distributions (generalization gradients) represented systematic distortions from the previously established English distributions. There were two types of distortions: of the slope of the gradients (increasing or decreasing the determinacy of the color names), and of the location on the physical continuum (making colors which in English have no specific names a perfect example of a color category in the nonsense language). Lenneberg found that if the tact generalization gradients ("reference relationships") of English were undistorted in the second language, the color words in that language were learned most easily; if only the locations of the gradients on the stimulus continuum were distorted, a small impairment in learning was observed. Learning the color names was most retarded when the shapes of the English color gradients were distorted in the second language to yield gradually rising slopes, that is, less well-defined boundaries between stimulus classes.

7. Mands

A mand is a verbal operant in which a response of a given form is reinforced by a characteristic consequence. The major variables that control a mand are the level of deprivation or of aversive stimulation applied to the speaker and his reinforcement history for similar responses. Questions, commands, entreaties, and the like are characterized by this kind of functional control (vide Skinner, 1957). Mands are difficult to condition in a self-instruction setting because an active listener, required to present appropriate reinforcement, is not available. It is unwieldy, if not impossible, to arrange characteristic reinforcing consequences, such as providing the time if the student should inquire, or complying with a command such as "stop," or granting an entreaty, and so forth. As was the case with tacts, the functional class of mands is largely neglected in second-language programming in the hope that, when deprivation or aversive stimulation occasions a mand, the transfer of certain intraverbals will take place. As is the case with tacts, this slight of mand presumably depends on mediated generalization, which may be a mixed blessing, facilitating the thematic accuracy of the response but degrading its formal properties. Although several second-language programs, notably those of Lermontoff (1961), Morton (1960), and Rocklyn and Moren (1962), indirectly condition mands by conditioning intraverbals of appropriate form (e.g., "ask for an apple in Spanish"), this may not be the only way of proceeding. An audio-lingual program could be so carefully sequenced that the occurrence of particular mands by the student was predictable and, therefore, the characteristic (verbal) reinforcers could be prerecorded on the master tape.

Transfer within and between repertoires

The concept of transfer has played a large role in this disquisition on programmed learning of a second language; it has turned up as an observation, an explanation, and an assumption at many points, so it may be well to summarize the sources of transfer available to, and often relied upon by, the second-language programmer. In echoic behavior and dictation, transfer of first-language auditory discriminations will occur, enhancing the acquisition of discriminations between sound classes but impeding the learning of equivalences (generalization) within these classes. The typical effect is over-discrimination and the emphasis in conditioning must be on the equivalences among discriminably different sounds. In textual behavior and transcription, transfer of first-language orthographic discriminations will occur when the second language employs an orthography that overlaps with that of English.

In echoic and textual behaviors, transfer of the differentiated vocal repertoire of the student's first language will occur to the degree that the two languages share a similar sound system or orthography. This greatly expedites the acquisition of a rough approximation to the desired repertoire, but it impedes the development of accurate pronunciation to the degree that the first- and second-language vocal repertoires differ. The consequences of transfer are similar for the differentiated manual repertoire involved in dictation and transcription.

A textual response will transfer to an echoic repertoire if the relevant auditory discriminations have been conditioned, and the converse can occur if the necessary orthographic discriminations have been learned in the first- or second-language. Similarly, transfer may occur between dictation and transcription if the relevant discriminations have been acquired. In complementary fashion, the student with an echoic repertoire should more readily acquire dictation when taught how to write the orthography than one who does not possess that repertoire; a similar transfer of discriminations may be expected to occur between transcription and textual behavior. These transfer effects are most obvious when the second-language orthography differs from that of the student's native language.

Within the thematic repertoires, an intraverbal response may transfer to the status of a tact in the second language; this requires non-verbal stimuli that evoke tacts in the student's native language which are linked intraverbally to responses in the second language. Intraverbal responses may similarly transfer to the status of mands in the second language when deprivation or aversive stimulation makes a mand likely and first-language responses intervene.

Transfer also takes place between the formal and thematic repertoires. A spoken response, whose form was conditioned during the acquisition of a purely echoic or textual repertoire, can take on the functional properties of an intraverbal through the replacement of one kind of stimulus control (formal) for another (thematic). This intraverbal response may then become a tact or a mand, in the manner described above. Similarly, written responses, learned as a part of the formal repertoires dictation or tran-

scription, may transfer to the thematic repertoires under suitable conditioning procedures.

Finally, I should note that the formal and thematic repertoires acquired in the language laboratory and in the language classroom do transfer, in some measure, to the second-language community.

References

- Anastasi, A. Psychological Testing. New York: Macmillan, 1961.
- Asher, J. J. Sensory interrelationships in the automated teaching of foreign languages. Percept. mot. Skills, 1962, 14, 38.
- Barcus, D. L., Hayman, J. L., and Johnson, J. T. Development of programmed learning materials for use with televised Spanish instruction. Unpublished report, Inst. for Commun. Res., Stanford Univ., Stanford, 1961.
- Berko, J. The child's learning of English morphology. Word, 1958, 14, 150-177.
- Blyth, J. W. (Ed.) The Hamilton College experiment in programmed learning. Unpublished report, Hamilton College, Sept. 1961.
- Brethower, D. M. On linguists and psychologists in second language learning, or the necessity of strange bedfellows. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.
- Brooks, N. Language and language learning; theory and practice. New York: Harcourt Brace, 1959.
- Brown, R. and Berko, J. Word association and the acquisition of grammar. Child Developm., 1960, 31, 1-14.
- Brown, R. and Fraser, C. The acquisition of syntax. In C. N. Cofer (Ed.) Verbal learning and verbal behavior. New York: McGraw-Hill, in press.

Carroll, J. B. Wanted: a research basis for educational policy on foreign language teaching. Harvard Educ. Rev., 1960, 128-140.

Carroll, J. B. An application of psycholinguistics in language teaching: an audio-visual instructional device. Paper read at Twelfth Ann. Round Table Meeting on Linguistics and Language Studies, Georgetown Univ., Washington, D. C., 1961.

Carroll, J. B. Research on teaching foreign languages. In (Ed.) Handbook of research on teaching. New York: Rand McNally, 1962. (a)

Carroll, J. B. What the foreign language teacher trainer or supervisor should know about programmed instruction in the foreign language field. Paper read at Foreign Language Teacher Training Seminar, Univ. of Washington, Seattle, 1962. (b)

Cooper, F. S., Liberman, A. M., and Borst, J. The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proc. Natl. Acad. Sci., 1951, 37, 318-325.

Cross, D. V. Metric properties of multidimensional stimulus control. Unpublished report, Ann Arbor, Univ. of Michigan, 04411-3-F, 1963.

Cross, D. V. and Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency and topography in auditory generalization. J. exp. anal. Behav., 1962, 5, 487-496.

Fant, C. G. M. Modern instruments and methods for acoustic studies of speech. Acta Polytech. Scand., 1958, 1, 1-81.

- Ferster, C. B. and Sapon, S. M. An application of recent developments in psychology to the teaching of German. In A. A. Lumsdaine and R. Glaser (Eds.), Teaching machines and programmed learning. Washington, D. C.: Nat. Educ. Assn., 1960.
- Fiks, A. I. and Garvey, C. J. Construction and evaluation of a short, automated, Vietnamese course. Unpublished report, Washington, D. C.: Human. Resour. Res. Off., Jan. 1963.
- Garvey, C. J. and Clark, J. W. Investigation of the procedures for limited tactical language training in Mandarin Chinese. Unpublished report, Washington, D. C.: Human. Resour. Res. Off., Sept. 1962.
- Geis, G. Possible reinforcements in a program. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.
- Gilbert, T. F. Mathetics: the technology of education. J. Mathetics, 1962, 1, 7-73.
- Gilpin, J. A versatile apparatus for audio self-instruction (the "Poly-audio"). In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.
- Gallanter, E. Contemporary Psychophysics Ch. 2 in T. M. Newcombe (Ed.), New directions in psychology, New York: Holt, Rinehart, and Winston, 1962.
- Hammond, P. C. The elementary biblical Hebrew program of Princeton Theological Seminary. Unpublished report, Princeton Univ., 1962.

- Harms, L. S. A self-instructional program for phonetic transcription. Unpublished report, Louisiana State Univ., Feb. 1962.
- Harris, C. C. An experience in self-teaching. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.
- Hayes, A. S. Language laboratory facilities. Washington, D. C.: U. S. Govt. Printing office, 1963.
- Hayes, A. S., Lane, H. L., Mueller, T., and Sweet, W. E. In W. E. Boltiglia, (Ed.) Current issues in language teaching - reports of the working committees, 1962, Northeast Conference on the Teaching of Foreign Languages. Manchester, N. H.: Lew Cummings, 1962.
- Holland, J. G. Teaching psychology by a teaching machine program. Amer. Psychol., 1959, 14, 414.
- Holland, A. L. and Matthews, J. Application of teaching machine concepts to speech pathology and audiology. Amer. Sp. Hear. Assn., 1963, 474-482.
- Israel, M. L. Variably blurred prompting: I. Methodology and application to the analysis of paired associate learning. J. Psychol., 1960, 50, 43-52.
- Keller, F. S. and Schoenfeld, W. N. Principles of psychology. New York: Appleton-Century-Crofts, 1950.
- Lane, H. L. On the discontinuity of speech discrimination learning in human adults. Unpublished report, Ann Arbor: Univ. of Michigan 04411-1-P, 1961..(a)

- Lane, H. L. Teaching machines and programmed learning. Harvard Educ. Rev., 1961, 31, 469-476. (b)
- Lane, H. L. Some differences between first and second language learning. Lang. Learn., 1962, 12, 1-14. (a)
- Lane, H. L. Psychophysical parameters of vowel perception. Psychol. Monogr., 1962, 76, 44. Whole no. 563. (b)
- Lane, H. L. Experimentation in the language classroom: guidelines and suggested procedures for the classroom teacher. Lang. Learn., 1962, 12, 115-121. (c)
- Lane, H. L. The motor theory of speech perception: a critical review. Unpublished report, Ann Arbor: Univ. of Michigan, 05613-1-P, 1963. (a)
- Lane, H. L. Differential reinforcement of vocal duration. J. exp. anal. Behav., 1963, in press. (b)
- Lane, H. L. Specifications for an auto-instructional device to shape the prosodic features of speech. Unpublished report, Ann Arbor: Univ. of Michigan, 04411-3-F, 1963. (c)
- Lane, H. L. and Moore, D. J. Reconditioning a consonant discrimination in an aphasic: an experimental case history. J. sp. hear. Disord., 1962, 27, 232-243.
- Lane, H. L. and Schneider, B. A. Methods for self-shaping echoic behavior. Mod. Lang. J., 1963, 47, 154-160.
- Lane, H. L. and Shinkman, P. G. Methods and findings in an analysis of a vocal operant. J. exp. anal. Behav., 1963, 6, 179-188.
- Lawrence, W. The synthesis of speech from signals which have a low information rate. In W. Jackson (Ed.), Communication theory. New York:

Academic Press, 1953.

Lenneberg, E. H. A probabilistic approach to language learning. Behav. Sci., 1957, 2, 1-12.

Lermontoff, A. Some aspects of programming a direct-method French course. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. of Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.

Lieberman, A. M. Some results of research on speech perception. J. acoust. Soc. Amer., 1957, 29, 117-123.

Lieberman, A. M., Harris, Katherine S., Kinney, JoAnn and Lane, H. L. The discrimination of relative onset-time of the components of certain speech and non-speech patterns. J. exp. Psychol., 1961, 61, 379-388.

Lieberman, A. M., Harris, Katherine S., Elmas, P., Lisker, L. and Bastian, J. An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. Lang. and Speech, 1961, 4, 175-195.

Licklider, J. C. R. Preliminary experiments in computer-aided teaching. In J. E. Coulsen (Ed.) Programmed learning and computer based instruction. New York: Wiley, 1962.

Licklider, J. C. R. and Miller, G. A. The perception of speech. Ch. 26— in S. S. Stevens (Ed.) Handbook of experimental psychology. New York: John Wiley, 1951.

Liedke, O. K. The programmed course in Elementary German. In J. W. Blyth (Ed.) The Hamilton College Experiment in programmed learning. Unpublished report, Hamilton College, Sept. 1961. App. II.

- Lisker, L., Cooper, F. S. and Liberman, A. M. The uses of experiment in language description. Word, 1962, 18, 82-106.
- Marty, F. Programming a basic foreign language course; prospects for self-instruction. Roanoke, Va.: Audio-Visual Publications, 1962.
- Mednick, S. A. and Freedman, J. L. Stimulus generalization. Psychol. Bull., 1960, 57, 169-200.
- Mildengerger, K. W. Problems, perspectives, and projections. Int. J. Amer. Ling., 1962, 28, 168-172.
- Miller, G. A. Free recall of redundant strings of letters. J. exp. Psychol., 1958, 56, 484-491.
- Miller, G. A. Some psychological studies of grammar. Amer. Psychol., 1962, 748-762.
- Miller, G. A. and Selfridge, J. A. Verbal context and the recall of meaningful material. Amer. J. Psychol., 1953, 63, 176-185.
- Morand, M. Report on the programming of French. In J. W. Blyth (Ed.) The Hamilton College Experiment in programmed learning. Unpublished report, Hamilton College, Sept. 1961. App. I.
- Morton, F. R. The language laboratory as a teaching machine. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. I, 1960.
- Morton, F. R. (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961. (a)
- Morton, F. R. (Ed.) Proceedings of the First Conference on the Role of

- the Language Laboratory in the Secondary School. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. IV, 1961. (b)
- Mostofsky, D. (Ed.) Stimulus generalization conference. Unpublished report, Boston Univ., 1963.
- Mueller, K. A. Experimentation and research in the development of modern foreign language materials and teaching methods. Int. J. Amer. Ling., 1962, 28, 92-104.
- Mueller, T. Programming morphemic structures: the concept of minute steps. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961. (a)
- Mueller, T. Revising Phase 1 of the programmed French Course, a work paper. Unpublished report, Univ. of Florida, 1962.
- Nostrand, H. (Ed.) Research on language teaching. Seattle: Univ. of Washington Press, 1962.
- Pimsleur, P. (Ed.) Report of the conference on psychological experiments related to second-language learning. Unpublished report, Language Development Program, U. S. Office of Education, Jan. 1960.
- Pimsleur, P. Discrimination training in the teaching of French pronunciation. Mod. Lang. J., 1963, 47, 199-203.
- Pimsleur, P. and Bonkowski, R. J. Transfer of verbal material across sense modalities. J. educ. Psychol., 1961, 52, 104-107.
- Pimsleur, P., Mace, L., and Keislar, E. Preliminary discrimination training

in the teaching of French pronunciation. Unpublished report, U. S. Office of Education, Contr. SAE 8950, 1961.

Porter, D. A report on the use of audio-visual aids in foreign language teaching. Unpublished report, Harvard Univ., April 1958.

Porter, D. A report on instructional devices in foreign language teaching. In A. A. Lumsdaine and R. Glaser (Ed.) Teaching machines and programmed learning. Washington, D. C.: Natl. Educ. Assn., 1960.

Reiff, D. G. Molecular analysis of Spanish phonology. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.

Rocklyn, E. H. Problems in programming an intensive aural-oral language course. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.

Rocklyn, E. H. and Clark, J. W. Construction, programming, and evaluation of a limited tactical language course. Unpublished report, Washington, D. C.: Human. Resour. Res. Off., Sept. 1962.

Rocklyn, E. H. and Moren, R. I. A feasibility study of a special machine-taught oral-aural Russian-language course. In S. Margulies and L. D. Eigen (Eds.) Applied programmed instruction. New York: Wiley, 1962.

Rocklyn, E. H., Moren, R. I., and Zinovieff, A. Development and evaluation of training methods for the rapid acquisition of language skills. Unpublished report, Washington, D. C.: Human, Resour. Res. Off., 1962.

- Saltzman, I. J. Techniques used in the construction of a completely self-instructional, one semester, modern college course in Russian. Unpublished address at Third Lang. Lab. Conf., Indiana Univ., Bloomington, March 1962.
- Sapon, S. M. An application of psychological theory to pronunciation problems in second-language learning. Mod. Lang. J., 1952, 36, 111-114.
- Sapon, S.M. Report on the use of a programmed course in Spanish. Unpublished report, Ohio State Univ., Columbus, 1961. (a)
- Sapon, S.M. Some reflections on models of linguistic structure and language programming. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961. (b)
- Sapon, S.M. Personal communication, 1963.
- Sapon, S. M. and Carroll, J. B. Discriminative perception of speech sounds as a function of native language. General Linguistics, 1958, 3, 62-72.
- Saporta, S. (Ed.) Psycholinguistics. New York: Holt, Rinehart and Winston, 1961.
- Saporta, S. Evaluation of three grammatical models in the teaching of foreign languages. In National Defense Language Development Research and Studies. Washington, D. C.: U. S. Govt. Print. Off. OE 12014-62, 1963.
- Schaefer, H. E. A. Poe as a reinforcer. Psychol. Rep., 1961, 8, 398.
- Scherer, G. A. C. and Wertheimer, M. The German teaching experiment at the Univ. of Colorado. German Quart., 1962, 35, 298-308.

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Special problems in programming language instruction for teaching machines. Int. J. Amer. Ling., 1960, 26, 167-174.

Stevens, S. S. Mathematics, measurement, and psychophysics. Ch. 1—in S. S. Stevens, (Ed.) Handbook of experimental psychology. New York: John Wiley, 1951.

Suppes, P., Crothers, E., and Weir, Ruth. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. Unpublished report, Stanford: Stanford Univ. Inst. for Math. Studies Soc. Sci., No. 51, Dec. 1962.

Suppes, P., Crothers, E., Weir, Ruth, and Trager, E. Some quantitative studies of Russian consonant phoneme discrimination. Unpublished report, Stanford: Stanford Univ. Inst. for Math. Studies Soc. Sci., No. 49, Sept. 1962.

Sweet, R. J. Programming control to elicit maximum effectiveness in pronunciation. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.

Sweet, W. E. The choice of syntactical and morphological discriminations in an elementary Latin program. In F. R. Morton (Ed.) Programming of audio-lingual language skills for self-instructional presentation. Ann Arbor: Publ. Univ. of Michigan Lang. Lab. Series Preprints and Reprints, Vol. VI, 1961.

Sweet, W. E. Outline and description of proposed Latin programs. Unpublished report, Univ. of Michigan, Sept. 1962.

Swets, J. A., Millman, S. H., Fletcher, W. E., and Green, D. M. Learning to identify nonverbal sounds: an application of a computer as a teaching machine. J. Acoust. Soc. Amer., 1962, 34, 928-935.

Uttal, W. R. On conversational interaction. In J. E. Coulsen (Ed.) Programmed learning and computer based instruction. New York: Wiley, 1962.

Valdman, A. The implementation and evaluation of a multiple-credit intensive audio-lingual approach in elementary French. Fourth and Sixth Progress reports, U. S. Office of Education, Contr. SAE 9498, 1962.

Wertheimer, M. and Scherer, G. A. C. The audiolingual teaching of German: a psycholinguistic experiment. Paper read at Psychonomic Soc., St. Louis, 1962.

Footnotes

1. The preparation of this article and certain of the experiments reported herein were supported in part by a contract with the Language Development Section, U. S. Office of Education. I am indebted to F. R. Morton and A. B. Gaarder who inspired and financed, respectively, my interest in second-language learning, and to D. M. Brethower, G. L. Geis, R. Glaser, and S. Sapon for helpful suggestions concerning this article.

2. See Carroll (1962a), Nostrand (1962) and Pimsleur (1960).

3. For example, see the review by Mednick and Freedman (1960) and studies reported in Mostofsky (1963).

4. Profound apologies to my secretary.

5. Additional contrasts, cited earlier, are the minimal units and their extension to new stimuli that characterize the formal but not the thematic repertoires. We may now qualify this distinction and note that thematic repertoires partake of both properties to a limited degree. Prefixes, suffixes, and roots approximate minimal thematic units. Extension to new stimuli is feasible for sets such as: concur, recur, incur, etc. Larger thematic units may also be extended; the student who knows the meaning of "elephant" has some notion, however crude, of the meaning of "elephantiasis," and so forth.

Figure captions

Fig. 1. Discrimination learning of Spanish phonemes with an audio-lingual program. The average number of frames in which three subjects reached criterion within one through 17 trials. (From Lane, 1961a.)

Fig. 2. Discrimination learning with an audio-lingual program. For each of 14 frames, the number of errors made during discrimination training is plotted against the percent increase in correct responses from a pre-test to a post-test, in which the program was administered without reinforcement. (From Lane, 1961a.)

Fig. 3. Proportion of errors on each day of discrimination training with Russian vowel phonemes. One group of 44 subjects ("R") received a fixed set of items in random order; a second group of 54 subjects ("P") received progressively longer and more difficult items. (From Suppes, Crothers, and Weir, 1962.)

Fig. 4. Learning to identify multidimensional nonverbal sounds with a computer-based program. The effects of two types of response, two types of feedback, conditional and unconditional selection of successive items, and three levels of reinforcement probability, on percent correct identification. (From Swets et al., 1962).

Fig. 5. Generalization gradients for pure tone intensity. The conditional probabilities were estimated from the total number of /ka/ re-

sponses (R_1) and /ti/ responses (R_2) emitted in ten presentations of each stimulus intensity to each of 20 subjects. The total latency (hexagons) at each intensity is the unweighted mean of the average latency of responding by each of 20 subjects. (From Cross and Lane, 1962).

Fig. 6. Generalization gradients for synthetic speech stimuli, ranging from /do/ to /to/. The conditional probability and average latency of the /do/ response in ten presentations of each stimulus to each of six subjects. (From Cross and Lane, 1962.)

Fig. 7. Differential reinforcement of vocal duration with one subject; shaping longer and then shorter durations. The relation between criterion response duration for differential reinforcement and the mean duration of unreinforced responses (inverted triangles), non-terminal reinforced responses (open triangles) and the ten terminal reinforced responses (filled triangles) in each phase. (From Lane, 1963b.)

Fig. 8. An audio-visual device for self-instruction in language learning. A sample film loop is resting above the projection screen, a sample magnetic tape cartridge along side the screen. To the left of the screen are a loudspeaker and a box containing relay circuitry. In the foreground is the control panel operated by the student and a write-in paper tape. (From Carroll, 1961.)

SUMMARY TABLE SECOND-LANGUAGE PROGRAMS
Available and in Preparation

Authors	Language	Behaviors Receiving Major Emphasis				Special Presentation Device	Approximate Hours to Completion	Location of Research	Comments and Source S = Study discussed here E = Evaluation data reported
		Echoic	Textual	Transcription	Dictation	Intraverbal	Tact	Mands	
Carroll	Chinese	x	x	x	x	x	x	x	Harvard Carroll, 1961 S
Ciou & Peterson	Chinese	x			x	x	x	yes	Univ. of Michigan 200
Garvey & Clark Chinese		x			x		x	tape recorder	
Barroughs	French	x			x			tape recorder 14-16	Encyclopedia Britannica Films
Mueller	French	x			x	x	x	yes	Univ. of Michigan 250
Hanzell	French	x	x	x	x	x	x	tape recorder year course	Univ. of Washington
Lemontoff	French	x	x		x	x	x	yes	Center for Programmed Instruction 200 S
Marty	French	x	x	x	x	x	x	yes	Hollins College year course Marty, 1962 S
Noraud	French	x	x	x	x	x	x	tape recorder year course	Hamilton College Blyth, 1961 S, E
Rickett & Dibols	French	x		x	x	x	x	18-23	General Program Teaching Corporation Carroll, 1962
Valdman	French	x			x	x	x	tape recorder year course	Indiana Valdman, 1962 S, E
Elert	German	x	x	x	x	x	?	?	Encyclopedia Britannica Films Carroll, 1962
Fenster & Spon	German			x	x			no	Harvard Fenster & Spon, 1960 S, E
Liedke	German	x	x	x		x	x	tape recorder year course	Hamilton College Blyth, 1961 S, E

Ventola & Wilson	German	x	x	no	17-30	TMI-Grollier	Carroll, 1962	
Bloom & Smith	Hebrew	x	x	no	15-25	TMI-Grollier	Carroll, 1962	
Hammond	Hebrew	x	x	no	48	Princeton	Hammond, 1962 Biblical material	E
Sweet	Latin	x	x	x	tape recorder two year course	Univer. of Michigan	Sweet, 1962	S
Thilson	Latin	x	x	no	undetermined	Center for Programmed Instruction		
Licklider	Russian	x	x	computer based	undetermined	Bolt, Beranek, & Newman	Licklider, 1962	S
Myers	Russian	x	x	yes	1/2 year course	Earlham College	Smith, 1962	S, E
Morton, Mayer, & Brethower	Russian	x	x	yes	275	Univer. of Michigan		
Rocklyn, Moren, & Zinovieff	Russian	x	x	tape recorder 140		George Washington University	Rocklyn, Moren, & Zinovieff, 1962 Limited tactical language	S
Saltzman	Russian	x	x	tape recorder 1/2 year course		Indiana	Saltzman, 1961	
Wilson & Ventola	Russian	x	x	no	18-22	TMI-Grollier	Carroll, 1962	
Barcus	Spanish	x	x	no	20	Denver Public Schools	Carroll, 1962	
Gilbert	Spanish	x	x	no		TOR Labs	Carroll, 1962	
Harris	Spanish	x	x	tape recorder 260		Louisiana State Univer.	Harris, 1961	S, E
Morton, Mayer, & Brethower	Spanish	x	x	yes	225	Univer. of Michigan	Morton, Mayer, & Brethower, 1960	S

Sapon	Spanish	x	x	x	x	x	x	x	x	tape recorder 50-85	Encyclopedia Britannica Films	Carroll, 1962	S, E
Sullivan	Spanish	x	x	x	x	x	x	x	?	tape recorder 40-45	Encyclopedia Britannica Films	Carroll, 1962	S, E
Universal Teaching Machines	Spanish	x	x	x	x	x	x	x	x	tape recorder 24-35	Universal Electric		
Wilson & Ventola	Spanish					x	x	x		17-30	TMI-Grolier	Carroll, 1962	
Morton, Mayer, Thai & Brethower		x				x	x	x	x	yes	Univer. of Michigan		
Fiks & Garvey	Vietnamese	x				x	x	x	x	tape recorder ?	George Washington University	Fiks & Garvey, 1963	S
Details not available for													
Bailey	Latin										Hollins College		
Newmark	French										Systems Development Corporation		
Unknown	Spanish					x				no	AVA Corporation		
Unknown	French										American Teaching Systems		
Unknown	Spanish										American Teaching Systems		
Unknown	German					x				no	AVTA Corporation		
Unknown	Hebrew										AVTA Corporation		
Unknown	Russian					x					AVTA Corporation		
Unknown	Russian										International Teaching Systems		
Koppitz	German					x				computer based	IBM	Uttal, 1962	
Allen	English										Thompson Ramo Woolridge		

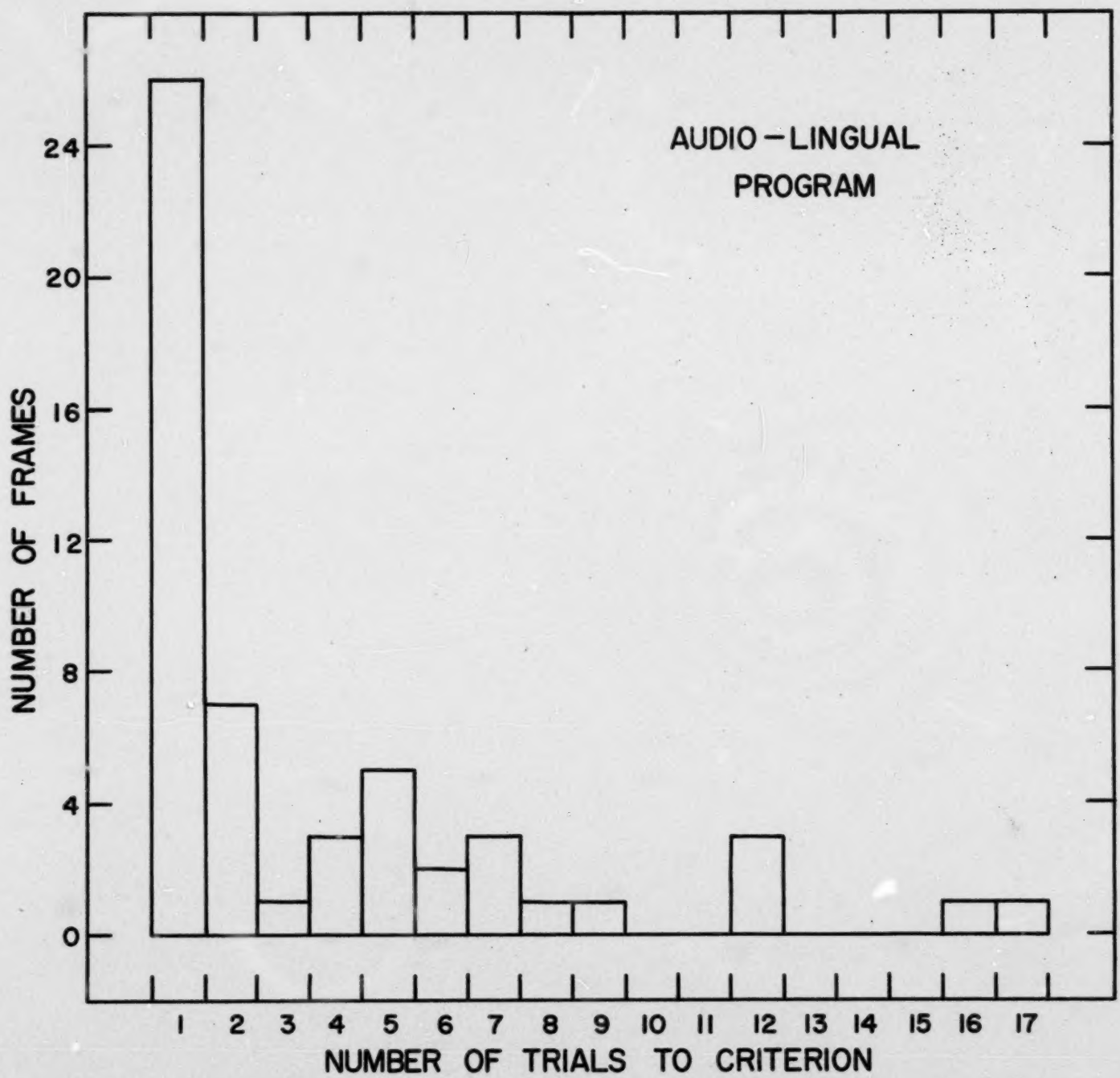


Fig. 1

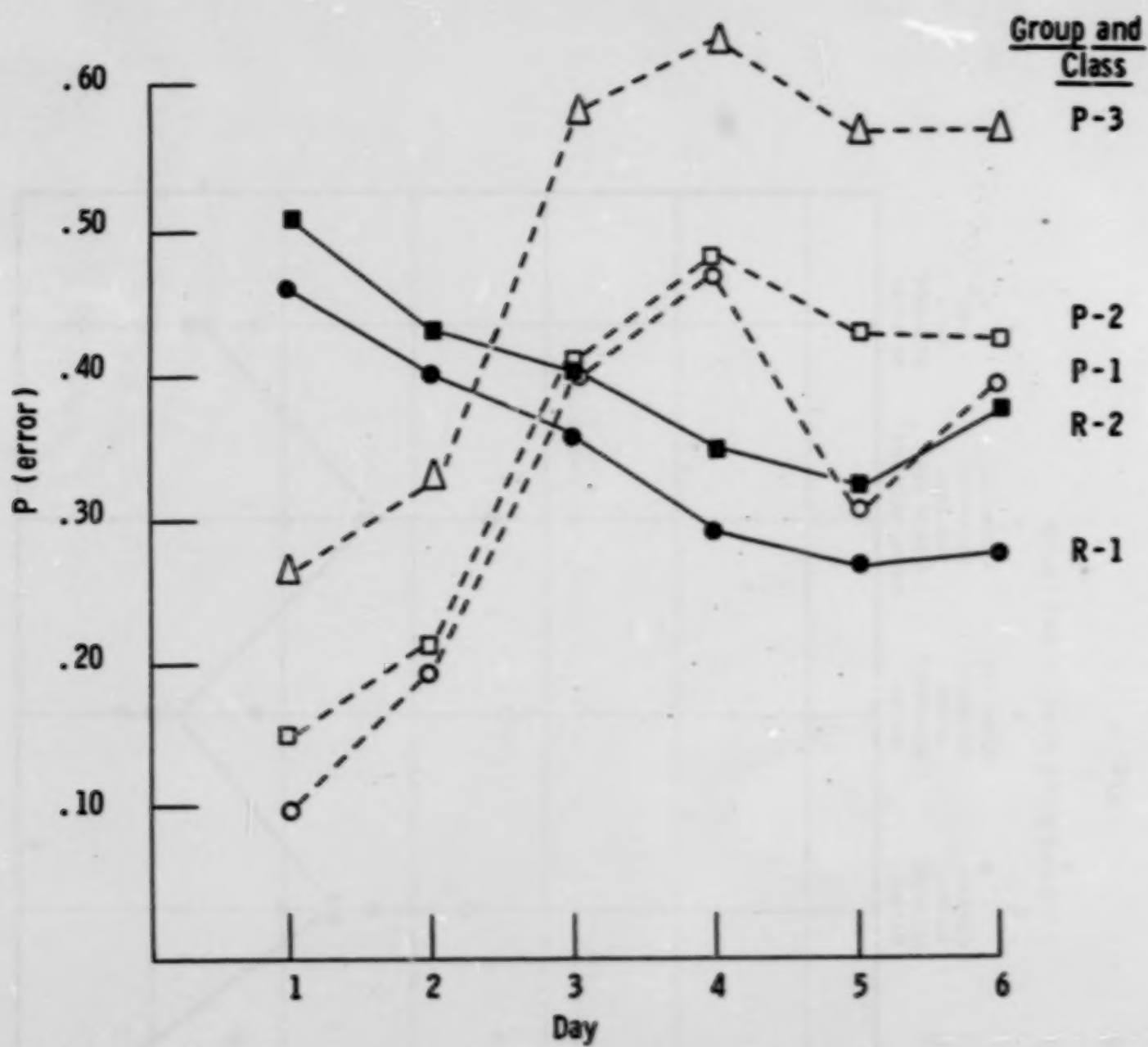


Fig. 3

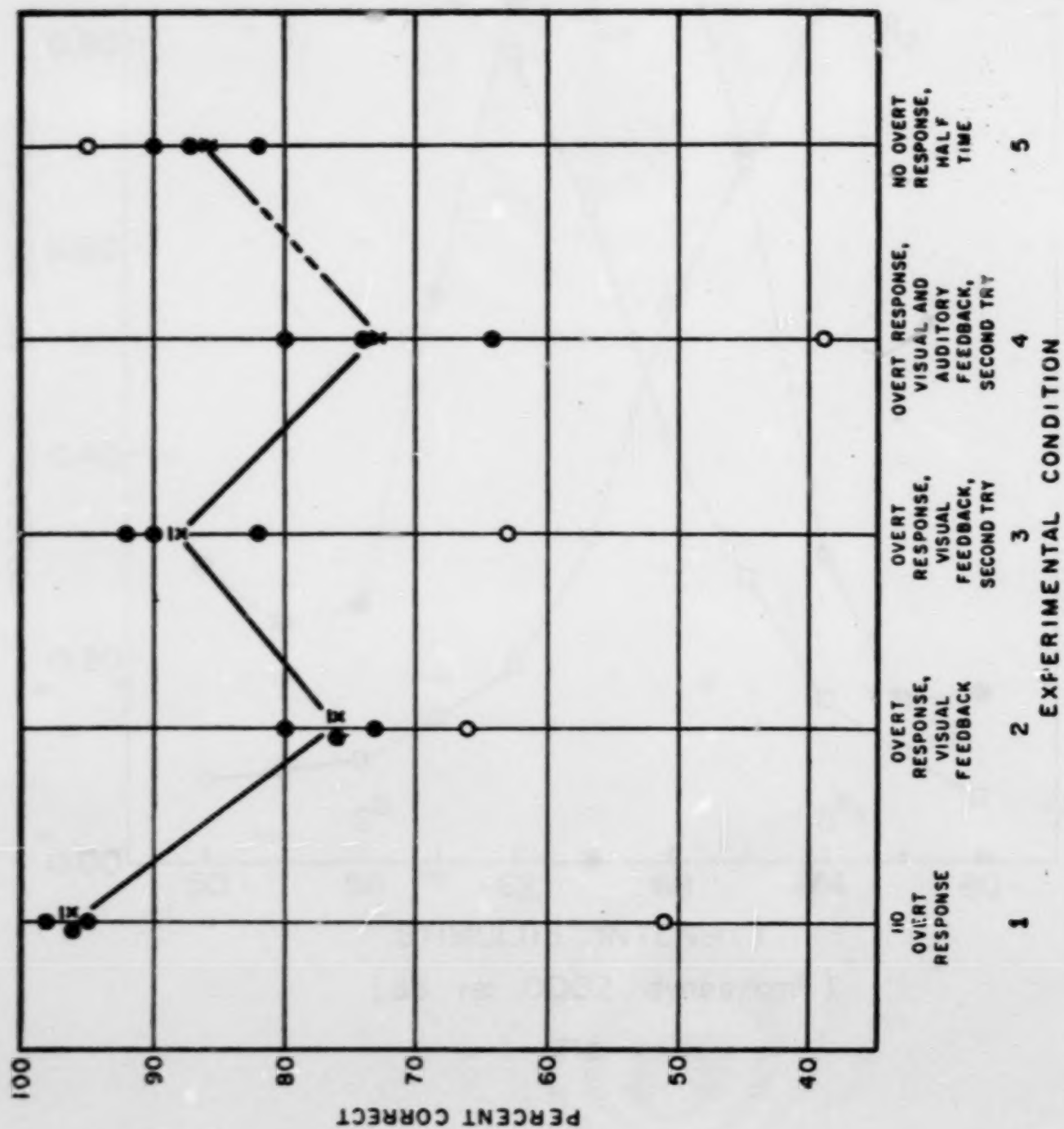


Fig. 4

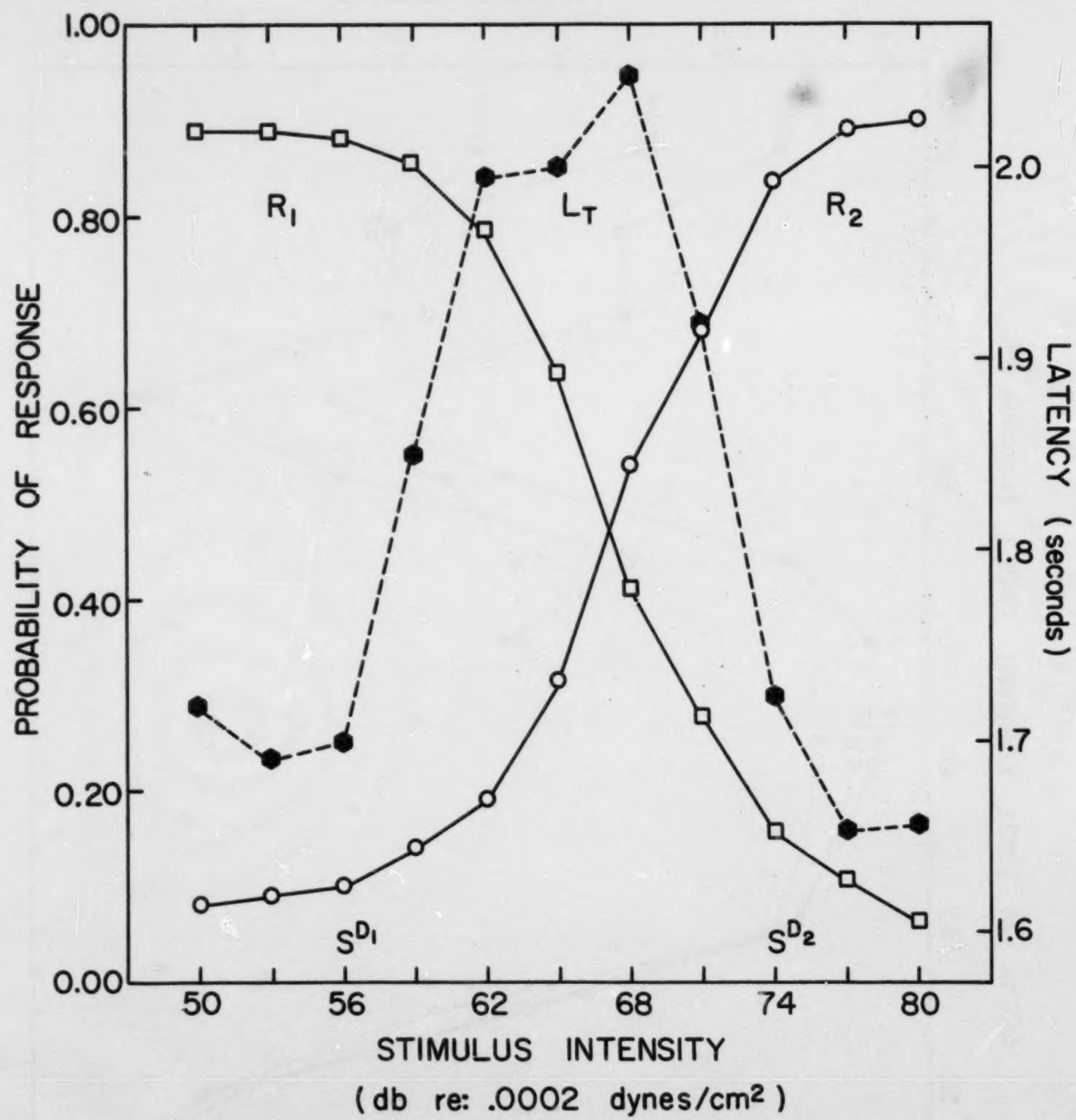


Fig. 5

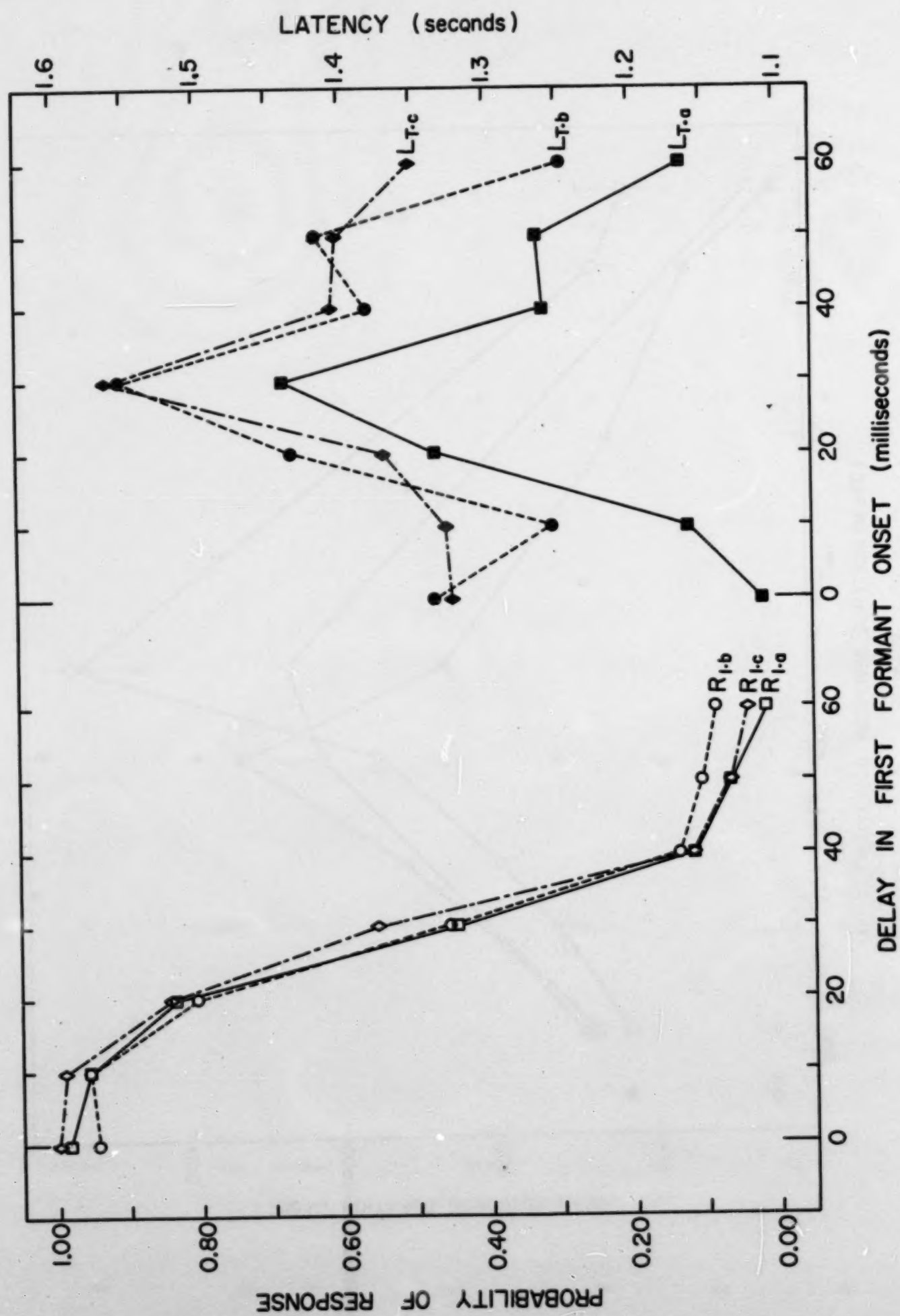


Fig. 6

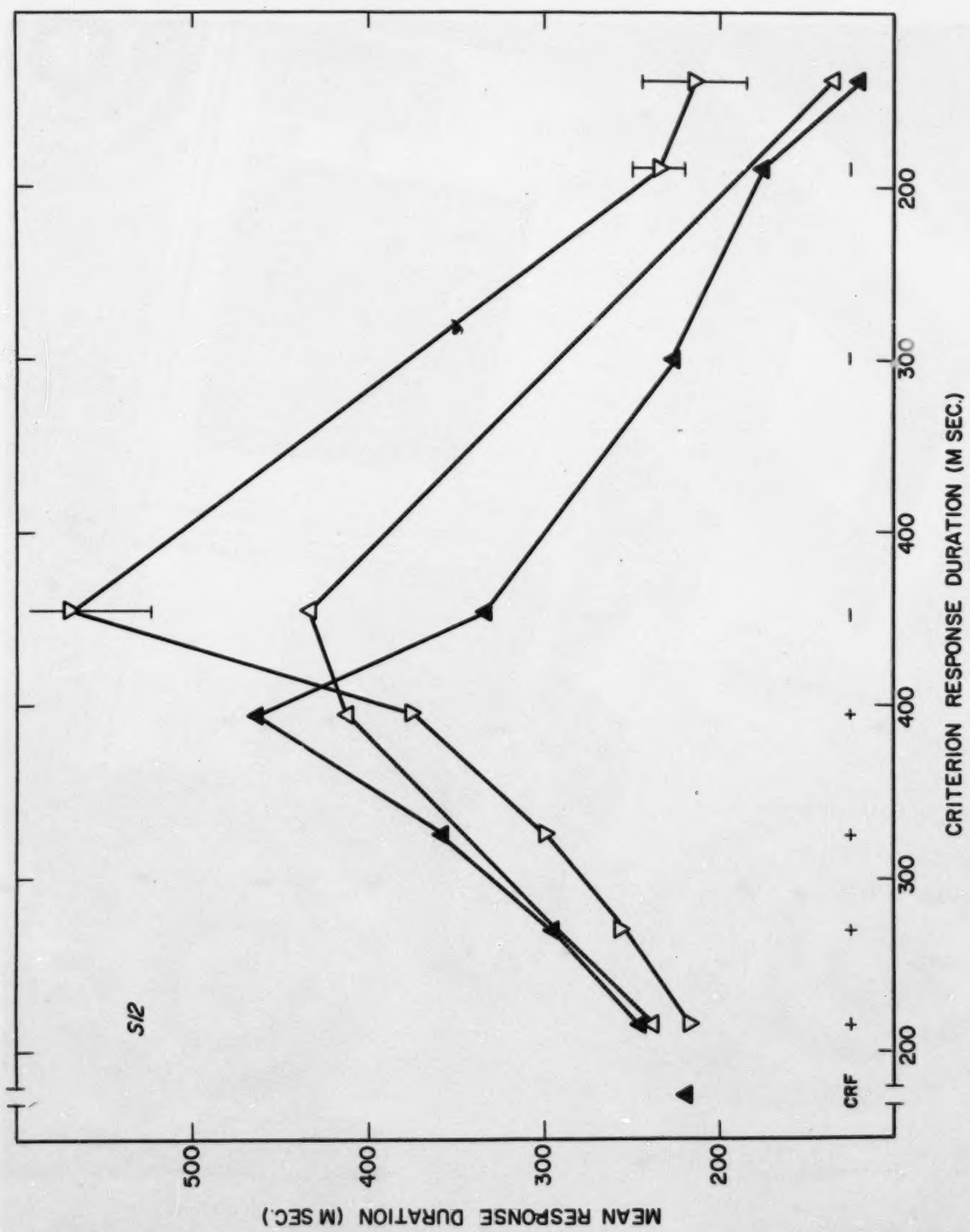


Fig. 7



Fig. 8

THE MOTOR THEORY OF SPEECH PERCEPTION: A CRITICAL REVIEW

Harlan Lane
The University of Michigan

The Motor Theory of Speech Perception: a Critical Review¹

Harlan Lane

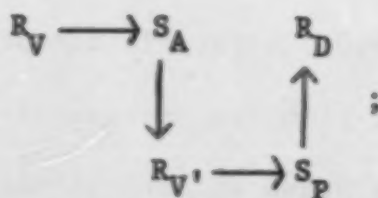
Behavior Analysis Laboratory

University of Michigan

The motor theory of speech perception maintains that "articulatory movements and their sensory effects mediate between the acoustic stimulus and the event we call perception" (Liberman, 1957). The observed sequence of events in a speech perception episode may be represented as:

$$R_V \longrightarrow S_A \longrightarrow R_D ;$$

a vocal response, R_V , generates an acoustic stimulus, S_A , which leads to a discriminative response, R_D . According to the motor theory of speech perception, the expanded sequence of events is:



the acoustic stimulus leads to a covert, articulatory response, R_V' , whose proprioceptive feedback, S_P , leads to the discriminative response.

The inference of mediating articulation in speech perception has appeared in several experimental contexts, including discussions of phonetics and phonemics, the discrimination of synthetic speech stimuli, the electromyography and acoustics of speech production, psychophysical functions for speech parameters, first-language acquisition and second-

language pedagogy. Although the hypothesis that speech is perceived by reference to articulation has not been elaborated into a formal theory, its recurrence and central role in these discussions suggests that a critical review is warranted. The purpose of this article is to bring together and examine the various experimental findings that have been interpreted as favoring a motor theory of speech perception, to present opposing findings and interpretations, and to show that, in the light of the available evidence, the inference that articulation mediates speech perception is unjustified.

The research findings that have been interpreted as favoring this theory have in common that the overt verbal behavior of the listener (R_D) in response to speech is more closely correlated with the patterns of articulation (R_V) that typically produce those speech signals than with the acoustic signals (S_A) themselves. From this observation it is inferred that corresponding patterns of articulation (R_V) actually mediate between the acoustic signals and the overt reports of the listener.² Since the speech signals generated by articulation are predictable from anatomical data (Fant, 1960), a correlation between the perceptual response and the acoustic stimulus necessarily implies a correlation between the perceptual response and the pattern of articulation that generated the stimulus. Therefore, the mere correlation of perception with articulation does not warrant the introduction of a mediating process in analyzing the listener's behavior in response to speech signals. In order to establish the usefulness of this inference, the experimental evidence must show that the perceived similarities and differences of a set of speech sounds correspond more closely to the articulatory than to the acoustic similarities and differences among

the sounds. In other words, R_V is inferred when there is a greater isomorphism between changes in R_V and R_D than between S_A and R_D .³

There are fundamentally three kinds of evidence, therefore, that have been viewed as favoring a motor theory of speech perception; these serve as the framework for the present discussion:

(1) When diverse perceptual responses are evoked by a set of acoustically similar speech signals, produced by diverse patterns of articulation.

(2) When similar perceptual responses are evoked by a set of acoustically diverse speech signals, produced by similar patterns of articulation.

(3) When both articulation and the acoustic signal change continuously over some range, and changes in the perceptual response correspond more closely to changes in articulation.

Evidence that favors a motor theory of speech perception

(1) Similar acoustic stimuli, diverse perceptual and articulatory responses.

Over the past few decades, scientists at the Haskins Laboratories have greatly advanced our understanding of the perception of speech by experiments involving synthetic speech-stimuli, generated by a device, called Pattern Playback, which permits careful control of stimulus parameters.⁴ Among the earliest studies whose findings are seen as favoring a motor theory of speech perception, is an investigation of the perception of unvoiced stop-consonants by Liberman, Delattre and Cooper (1952), employing the Pattern Playback to convert Handpainted

spectrograms into sound. Figure 1c shows one of the schematic syllables that was converted into sound, and Figs. 1a and 1b show the two acoustic

Insert Figure 1 about here

variables of the experiment: the frequency locus of a schematic stop and the formant frequencies of the following vowel. The 12 stops before each of the 7 vowels gave a set of 84 consonant-vowel (CV) syllables that were presented to SS for identification.

When the frequency locus of the schematic stop was held constant and the following vowel was varied over the seven pairs of formant frequencies, shifts in the modal consonant identification were obtained. For example, the schematic stop at 1440 cps was called /p/ (59 times out of 60) before /i/, but it was increasingly often called /k/ as the following vowel ranged from /i/ to /a/; it was called /k/ (57/60) before /a/ but increasingly often called /p/ as the following vowel ranged from /a/ to /u/; the same stop was called /p/ (42/50) before /u/.

In order to account for the diverse consonant identifications evoked by a set of invariant schematic stops, the investigators pointed to the diverse patterns of articulation necessary to generate the natural language correlates of their acoustic stimuli:

"With the mouth set to articulate /i/ or /u/, a speaker would presumably use a movement of the lips to produce a sound having roughly the same acoustic characteristics as our schematic stop at 1440 cps; for the vowel /a/, the same acoustic result would be most closely approximated by lowering the back of the tongue from a point of contact with the roof of the mouth." (p. 513)

The authors then suggest that these diverse patterns of articulation are operative during acoustic reception to translate acoustic similarity into perceptual diversity:

"We assume, then, that the sound patterns corresponding to the syllables pi (or pu) and ka set off the appropriate movements in the listener (that is, the articulatory movements which the listener would make in attempting to reproduce these acoustic patterns) or perhaps only the short-circuited neural equivalents of these movements, with the result that the initial schematic phonemes p and k, which can be entirely identical as acoustic stimuli, become clearly differentiated in proprioceptive terms and are therefore finally perceived as distinctly different sounds." (p. 514)

Several more recent experiments by Liberman and co-workers have also examined labelling responses to synthetic consonant-vowel pairs, and have provided support for the motor theory of speech perception. The five studies summarized below have in common that the acoustic parameters of the stimuli that controlled consonant identification all varied regularly over some range, while the parameters of the following vowel were held constant.

Liberman et al. (1957) synthesized a series of 14 speech stimuli, using the patterns shown in Fig. 2, which varied only in the direction

Insert Figure 2 about here

and extent of second-formant transition. When the auditory stimuli were presented in irregular order to a group of listeners for identification, the labelling functions that were obtained resembled the one shown

in Fig. 3 for a typical S. The relatively steep gradients that define

Insert Figure 3 about here

the phoneme boundaries /b/ - /d/ and /d/ - /g/, are seen by the authors as favoring the inference of mediating articulation. The articulatory patterns that, in natural language, generate the three phonemes may be viewed as diverse (i.e., discretely different): /b/ is a bilabial plosive, /d/ is alveolar, and /g/ velar. The gradients, which "partition" the stimulus continuum, reflect diverse labelling responses. However, the speech stimuli, drawn from an acoustic continuum, are considered similar. It is inferred, therefore, that diverse mediating responses "map" the stimulus continuum onto diverse perceptual responses.

A later study by Liberman et al. (1961a) provides additional evidence of this kind for the motor theory of speech perception. Seven synthetic speech stimuli were employed that varied only in the delay of the onset of the first formant relative to that of the second and third formants and that ranged perceptually from /do/ to /to/. The "cutback" in the first formant ranged from 0 to 60 msec in 10-msec steps. The labelling function for a typical S shows greater than 95 percent /do/ identification for stimuli 1 and 2 (= 5 percent /to/), 80 percent at stimulus 3, 14 percent at stimulus 4, 5 percent at stimulus 5, and no /do/ identifications at stimuli 6 and 7. "For all Ss, the phoneme boundary between /d/ and /t/ is quite sharp; in every case, a change of 10 msec in the first formant cutback is sufficient to shift the responses from 75 percent /d/ to 75 percent /t/" (p. 383). The near-dichotomy in labelling responses corresponds to the voiced-voiceless dichotomy in classifying the articulations that produce /d/ - /t/ in natural language.

The next study in this series, conducted by Liberman et al. (1961), employed the phonemic contrast rabid-rapid. Twelve speech stimuli were synthesized with intervocalic silent intervals that varied from 20 to 140 msec in steps of 10 msec. Figure 4 shows the relative frequency of /b/ and /p/ responses to these stimuli by 12 listeners. The investigators conclude from these data that, once again, "the phoneme boundary

Insert Figure 4 about here

is reasonably sharp" (p. 183). As in the prior study, the corresponding dichotomy in classifying articulation of the speech stimuli in natural language is voiced-voiceless.

The next pair of studies provides additional evidence of "categorical perception" of speech stimuli and also electromyographic data to confirm that the patterns of articulation that generate these stimuli are similarly categorical. Bastian, Eimas, and Liberman (1961) inserted silent intervals of varying lengths into natural recordings of the word slit and presented them to listeners for identification as either slit or split. "The perception of this continuum can be regarded as essentially categorical...The categorical perception of the consonants may be explicable in terms of the categorical nature of their articulatory gestures," (i.e., the absence or presence of the bilabial plosive).

Harris, Bastian and Liberman (1961) have shown that, when ss were instructed to mimic stimuli sampled from the slit-split continuum, the silent intervals in their responses fell into a bimodal frequency distribution; they were, in other words, similarly categorical. To determine whether there were incipient /p/ movements in the region of the phoneme boundary, electromyographic recordings were made during mimicry;

these were also entirely categorical: "either there was a normal burst of muscle potential at the lip (indicating a /p/ gesture) or there was not."

In each of the phoneme-labelling experiments described above, the investigators also presented their listeners with the synthetic speech stimuli arranged in ABX triads, in order to measure percent correct discrimination within pairs of stimuli drawn from the speech continuum. In accordance with the motor theory of speech perception, the experimenters expected to find that two stimuli drawn from the same phoneme category (defined by the labelling gradients) are less often correctly discriminated than two stimuli drawn from opposite sides of the phoneme boundary. The rationale is spelled out by Liberman et al. (1961):

"We believe that in the course of his long experience with language a speaker (and listener) learns to connect speech sounds with their appropriate articulations. In time, these articulatory movements and their sensory feedback (or, more likely, the corresponding neurological processes) become part of the perceiving process, mediating between the acoustic stimulus and its ultimate perception. When significant acoustic cues that occupy different positions along a single continuum are produced by essentially discontinuous articulations...the perception becomes discontinuous (i.e., categorical), and discrimination peaks develop at the phoneme boundary." (p. 177)

The experiments to be described, which examined ABX discrimination of synthetic speech stimuli, had two further objectives. The first was to see to what degree the data warrant the extreme assumption that

"the listener cannot hear any differences among these sounds beyond those that are revealed by his use of the phonemic labels," (Liberman, Dellatre, and Cooper, 1952, p. 359). This outcome may be expected if both the labelling responses and ABX discriminations are entirely dependent on the proprioceptive feedback from the mediating articulation hypothesized by the motor theory of speech perception. The contrary assumption, namely that the listener can discriminate among many more stimuli than he can identify absolutely, is in accord with psychophysical data on the perception of non-speech stimuli (see, for example, Miller, 1956).

The third and last objective of the discrimination studies that bears on the motor theory of speech perception was to assess the degree to which the discrimination functions reflected "acquired similarity," "acquired distinctiveness" or both:

"One possibility [acquired similarity] is that in the raw these speech stimuli were all as highly discriminable as the most discriminable pair, and that the effect of our linguistic experience has been to dull our sensitivity to the differences within the phoneme category...the contrary possibility is that our stimuli were originally as little discriminable as the least discriminable pair, and that our discriminations have been selectively sharpened as a result of our long experience with the language" (Liberman, 1957, p. 122).

The motor theory of speech perception is seen by Liberman as favoring the latter outcome:

"It is tempting to speculate that what is learned is

simply a connection between various acoustic stimuli and certain articulatory responses...We should suppose that the perceived difference between two relatively similar external stimuli could be increased if we could attach to those stimuli two very different mediating responses and hence gain the added distinctiveness of their very different proprioceptive returns" (p. 123).

Discriminations among synthetic speech stimuli were measured, in the first of the phoneme-labelling experiments described above (Liberman et al., 1957), by arranging pairs of the 14 stimuli in triads of the form ABA and ABB. Three subsets of all possible pairs of stimuli were employed: "one step" comparisons, involving pairs of adjacent stimuli on the acoustic continuum, "two step" comparisons, involving pairs of alternate stimuli, and "three step" comparisons. The discrimination functions obtained for a typical listener are shown in Fig. 5. Comparison

Insert Figure 5 about here

of these data with the corresponding labelling gradients in Fig. 3 shows that the listener discriminates more accurately between sounds that lie on opposite sides of the phoneme boundary than he does between sounds (separated by the same number of physical units) drawn from within one phoneme category.

In order to assess the degree to which the listener "cannot hear differences...beyond those revealed by the phoneme labels", a set of predicted discrimination functions was derived from the labelling gradients, following the basic rule that the greater the difference between two stimuli in the frequencies with which each was assigned a particular phoneme label, the greater their predicted discriminability.

Since the obtained discrimination functions correspond in form to the predicted functions, but lie somewhat above them, the authors conclude that listeners discriminate the speech stimuli largely to the extent that they label them differently, but they are able to extract some information in addition to that reflected in the labelling gradients (cf. Fig. 5).

In order to assess the degree to which the speech discrimination functions reflect acquired similarity, acquired distinctiveness, or both, it is necessary to have a baseline of discrimination which presumably reflects the discriminability of the speech stimuli "in the raw" — that is, before language learning has taken place. The desire for a baseline that, on the one hand, can be obtained from adults, who already have acquired speech discriminations, and, on the other hand, has some face validity as an index of discrimination in its pristine form, places two mutually incompatible requirements on the set of control stimuli used to obtain this baseline: first, the control stimuli should be comparable to the speech stimuli, varying in the same acoustic properties and with the same constant parameters, so that the difference in discriminability of the speech and control stimuli may not be attributed to a difference in the psychophysical task; second, the control stimuli must not evoke the speech discriminations under study, so that their discriminability may indeed serve as a baseline for assessing the effects of training in the speech discrimination.

Evidence for acquired distinctiveness of the speech sounds, which is interpreted as favoring the motor theory of speech perception, could not be obtained in the study of the phonemes /b/ - /d/ - /g/ described above, because a suitable set of control stimuli was not prepared. In

experiments with the contrasts do-to, rabid-rapid, and slit-split, summarized below, the obtained discrimination functions provided all three kinds of evidence favoring the motor theory of speech perception: increased discriminability at the phoneme boundaries, close correspondence between obtained and predicted discrimination functions, and much greater discriminability of the speech stimuli than their non-speech controls, showing acquired distinctiveness.

In the experiment employing speech stimuli that ranged perceptually from /do/ to /to/ (Liberman et al., 1961a), the stimuli were arranged in one-, two- and three-step comparisons in ABX triads of the form: ABA, ABB, BAA and BAB. The discrimination functions obtained from eleven listeners agreed well with those predicted from their labelling gradients. "Discrimination was better for a given stimulus difference when the two stimuli lay on opposite sides of a phoneme boundary than when the stimuli were within the same phoneme category" (p. 384).

In order to obtain a set of non-speech, control stimuli that differed with respect to cues comparable to those defining the speech continuum, the authors inverted their speech spectrograms before converting them to sound with the Pattern Playback.⁵ The control stimuli were prepared in ABX triads according to the same procedure followed for the speech stimuli and were presented a like number of times to the same listeners employed for the labelling and discrimination phases of the experiment. The discrimination functions that were obtained for the control stimuli did not depart appreciably from chance, they did not show any regularly occurring peaks, and they lay well below the speech-discrimination functions in all comparisons.

The authors conclude that, to the extent that the control stimuli provided a valid measure of the discriminability of the speech stimuli

prior to language learning, the speech discrimination functions reflect acquired distinctiveness to a large degree. Two other sources of evidence may be cited in support of this conclusion. Hirsch (1959) had listeners judge the temporal order of two pure tones of differing frequency. The difference in time of onset to give 75 percent correct judgment was about 17 msec. In comparison, the difference in time of onset of the first and second formants of the synthetic speech stimuli that gave 75 percent correct discrimination was about 12 msec. Thus, the peak in the /do/ - /to/ discrimination function again may be considered to reflect acquired distinctiveness. Lane and Moore (1962) obtained a "non-speech" control baseline for the /do/ - /to/ discrimination by measuring, with an aphasic subject, the discriminability of the same set of speech stimuli employed by Liberman et al. (1961a). Chance levels of discrimination (and a flat labelling gradient) were observed prior to discrimination training. Following training, a gradient of phoneme identification was obtained and ABX discrimination was well above chance levels in all three-step comparisons, peaking at the phoneme boundary.

In the experiment employing 12 speech stimuli that ranged perceptually from rabid to rapid (Liberman et al., 1961), the stimuli were arranged in two-step through nine-step comparisons in ABX triads. The discrimination functions obtained from 12 listeners, shown by the filled circles in Fig. 6, reveal a peak at the phoneme boundary (cf. Fig. 4)

Insert Figure 6 about here

and, in general, agree well with those functions predicted (unfilled circles) from the labelling gradients. The obtained discrimination

functions lie, at all points, somewhat above those predicted, suggesting that the listener can discriminate between two speech stimuli slightly more often than he can identify them absolutely.

The speech discrimination functions shown in Fig. 6 are not, however, unimodal. There is one peak in the region of 70 msec, corresponding to the phoneme boundary shown in Fig. 4, and a second peak in the region of 100 msec that does not correspond to a phoneme boundary. The authors considered that the stimuli with greater than 100-msec silent interval, labelled by their listeners as /p/, were "unnatural." They thought that this phoneme "might nevertheless be heard and articulated by our listeners almost as if it were a different speech entity." The listeners were then recalled and instructed to label the speech stimuli anew, this time with three possible responses: /b/, /p/, and */p/. The labelling function for the group of Ss then showed a third category, */p/, and hence a second "phoneme boundary": /p/ - */p/. The predicted discrimination functions shown in Fig. 6 were derived from this revised labelling gradient.

In order to assess the degree of acquired distinctiveness shown by the speech-discrimination functions, control stimuli were prepared by processing the speech stimuli so as to yield noise bursts with the same amplitude envelope and the corresponding silent intervals. These control stimuli were arranged and presented for discrimination measurement in the same manner as the speech stimuli. The discrimination of the control stimuli proved inferior to that of the speech stimuli in almost all comparisons, as shown in Fig. 6. "The entire learning effect consists of a sharpening of discrimination in the vicinity of the phoneme boundary" (p. 189). This conclusion agrees with that reached earlier in the experiment on discrimination of the phonemes /d/ - /t/.

In the experiment employing speech stimuli that ranged perceptually from slit to split, reported by Bastian, Eimas, and Liberman (1961), ABX discrimination was also measured and found to show peaks in the region of the phoneme boundary and, over all, a close correspondence to the discrimination functions predicted from the phoneme-labelling gradients. The acquired distinctiveness of the slit-split stimuli was not determined.

The following kinds of evidence, viewed by the experimenters as favoring the motor theory of speech perception, have been cited in this section: When the frequency locus of a schematic stop was held constant and the following vowel varied, consonant identification varied in a way that seemed to accord with the varied patterns of articulation necessary to produce the comparable speech stimuli in natural language. When synthetic speech stimuli were selected from an acoustic continuum that ranged perceptually from one phoneme to another, and these stimuli were presented to listeners for identification, the phoneme-labelling gradients were quite steep, suggesting categorical perception of the speech stimuli, mediated by the discretely-different articulations required to produce the two phonemes in natural language. When the synthetic speech stimuli employed for the labelling experiments were arranged in ABX triads for discrimination measurement, listeners more readily discriminated between stimuli drawn from opposite sides of the phoneme boundary than between stimuli drawn from the same phoneme category, suggesting that discretely-different mediating responses were involved in the former discriminations, enhancing discriminability by their very different proprioceptive returns. These discrimination functions accorded well with those predicted from the labelling gradients on the assumption that the listener could discriminate speech stimuli only to

the degree that he labelled them differently, suggesting that discrimination was controlled, not by the continuously-varying speech stimuli, but rather by the proprioceptive return from only two different mediating responses. Finally, when a baseline of discriminability was obtained for the speech stimuli by presenting appropriate control stimuli for discrimination measurement, the speech discriminations were found to be much more accurate in all comparisons, suggesting acquired distinctiveness of speech stimuli, brought about by the mediating articulations that these stimuli evoke.

(2) Diverse acoustic stimuli, similar perceptual and articulatory responses.

Liberman (1957) has put the question: "When articulation and sound wave go their separate ways, which way does the perception go?" His answer, favoring the motor theory of speech perception, was: "The perception always goes with articulation." Part of the support for this answer is provided by the experiments discussed above, in which diverse articulations produced similar acoustic stimuli that evoked diverse perceptual responses. Additional support for the answer is provided by the complementary findings, discussed below: similar articulations produce diverse acoustic stimuli that evoke similar perceptual responses.

The first evidence of this sort comes from a study of the perception of the stop consonants, discussed earlier (Liberman, Delattre, and Cooper, 1952). When the frequency locus of a schematic stop and the formant frequencies of a following vowel were both allowed to vary, the greatest number of /k/ identifications was not associated with any one schematic stop, but varied widely, depending on the following vowel; for example, a schematic stop at 2520 cps gave the most /k/ judgments if the following

vowel was /e/, but one at 710 cps gave the most /k/ judgments if the following vowel was /u/. The nearest spoken equivalents to the schematic /k/'s are all produced by similar articulatory movements (although the point of the tongue tends to be slightly back for back vowels and more forward for front vowels). "We may say then, in regard to schematic stops heard as k, that the perception reflects the basic similarity in the articulatory patterns, rather than the relatively gross differences among the acoustic stimuli" (p. 514).⁶

Additional evidence seen as favoring the motor theory of speech perception is provided by experiments on the perception of the stop consonant /g/ in initial position. Liberman et al. (1954) have shown that the direction and degree of second-formant transition serves as a cue for the perception of the stop consonants, and that the degree or rate of transition necessary for modal identification of a given stop consonant in a CV pair depends on the vowel that follows it. Therefore, the perception of the consonant remains invariant, from one CV pair to the next, despite large changes in the formant transitions. With respect to the stop consonant /d/, the investigators were able to associate the perceptual invariance with an acoustic invariance, as shown in Fig. 7:

Insert Figure 7 about here

different /d/ transitions before different vowels "seem to be coming from the same frequency position -- namely the /d/ locus at 1800 cps" (Liberman, 1957, p. 121). In the case of /d/, therefore, perception, articulation and the acoustic stimulus all "go" the same way. In the case of /g/, however, no acoustic invariance was found to be associated with the invariant perception (or with the similar articulations that produce

/g/ before the several vowels). Figure 7 shows that this consonant requires a large transition originating at about 3,000 cps when the following vowel is /i/, /e/, /ɛ/, or /a/. When the following vowel is /ɔ/, /o/, or /u/, however, the transition is, on the contrary, very small and the locus has shifted markedly.

"Here, the large and abrupt change in acoustic pattern contrasts with the comparatively small and continuous shifts in articulation...The identity of the perceived consonant remains the same throughout, and thus parallels the invariance of the articulation" (Cooper et al., 1958, p. 937).

Perhaps the most striking invariance in the perception of speech under wide changes in the acoustic stimulus is the constancy of vowel identification despite changes in the vowel formant frequencies, due to differences among speakers, dialects, and phonetic environments. If each vowel phone, uttered in various phonetic contexts by a speaker, is represented as a point on a vowel formant chart, whose coordinates are the first- and second-formant frequencies of the vowels, and if the set of points corresponding to each phoneme is considered to define a vowel region, then the regions of different vowels sometimes overlap and the regions of the same vowels sometimes fail to overlap, as shown in Fig. 8.

Insert Figure 8 about here

In order to account for the perceptual constancy of the vowel phoneme despite wide variations in its acoustic properties, Joos (1948) put forth, in a classic monograph, a motor theory of vowel perception. Joos notes, first of all, that it is possible to find some order among the vowel

regions solely on the basis of their acoustic parameters. This ordering may be described graphically as a vowel triangle⁷ imposed on the formant plot; its apices are the vowel regions /i/, /a/, and /u/. When such a reference system is drawn for the vowels of a particular speaker, most of the other vowels he produces lie, to a degree, on the borders of or within the triangle, and in relatively predictable positions. The triangle fit to the utterances of a particular speaker, however, differs from one speaker to the next. Joos notes that the between-speakers variance in vowel regions may be greatly reduced by "sliding" the several vowel triangles into congruency and he describes a statistical procedure for approximating this. Joos suggests that this procedure is a model of the process by which phonemic constancy is achieved in perception. He contends, furthermore, that this normalizing process in vowel perception is controlled not by the speech signals themselves but by their articulatory correlates:

"The brain, then, contains an equivalent of Fig. 29

[a schematic of the organs of articulation] and upon this map it can lay down an equivalent of Fig. 26 [a vowel quadrilateral with the coordinates: height and position of the tongue-hump during articulation], and if it has two of the latter it can shift and distort one of them to fit the other" (p. 63).

Joos suggests that the brain may derive two vowel quadrilaterals much in the manner now used in certain automatic speech recognition procedures. When someone says, "How do you do?", the listener registers the first vowel phone as this speaker's articulation of low-central, the last phone as high-back, and the median /y/ would do to determine high-front. "The pattern's corners are now located, and other phones

can be assumed to be spaced relative to them as they are generally spaced in this dialect" (p. 61).

The following kinds of evidence, viewed by the experimenters as favoring the motor theory of speech perception, have been cited in this section: When the frequency locus of a schematic stop-consonant and the formant frequencies of the following vowel were allowed to vary, the same consonant identification was evoked by diverse acoustic patterns, suggesting mediation by the similar articulatory responses required in natural language to produce the consonant before the several vowels. When a stop consonant was synthesized before each of a series of vowels, by varying the transition of the second formant preceding its steady state, the same consonant identification was evoked despite a discontinuity in the series of second-formant transitions, suggesting, once again, mediation by the similar articulatory responses required to produce the consonant before the several vowels in the series. Finally, when the formant frequencies of natural-language vowels vary, two distinct vowel-formant regions, corresponding to utterances of the same vowel by two different speakers, evoke the same vowel identification, while two overlapping regions, corresponding to utterances of different vowels by different speakers, evoke different identifications. These findings suggest that the listener normalizes the acoustic vowel system of each speaker in order to map it onto a fixed set of vowel identifications, and that this is accomplished in terms of the articulatory correlates of the vowels rather than in terms of their more diverse acoustic parameters.

- (3) Changing perception is correlated less with the acoustic stimuli, more with the patterns of articulation that produce those stimuli.

The four experiments summarized below have in common that the acoustic variable was the sound pressure level of the utterance and the perceptual response was some estimate of loudness or stress.

The suggestion that the perception of a supra-segmental feature of language such as stress may involve mediating articulation was perhaps first made by S. Jones (1932) who wrote:

"Accent is sui generis depending for its perception on the kinesthetic sense...The listener refers what he hears to how he would say it. Thus he translates exteroceptor into proprioceptor sensations, the kinesthetic memory serving as stimulus."

In a study conducted by Ladefoged, Draper, and Whitteridge, (1958), the sentence: "The old man doddered along the road" was uttered by one speaker while action potentials of the internal intercostal muscles were measured. The authors report that listeners' estimates of the stress loci in the sentence corresponded to the temporal loci of heightened muscular activity. They suggest that statements about stress "are usually best regarded as statements about the speaker's muscular behavior (or about the actions of the listener's muscles which would have to be made in order to produce similar sounds)" (p. 9).

In a subsequent study by Draper, Ladefoged and Whitteridge (1959), speakers were instructed to count to twenty, making the words equally loud. The authors report that the peak intensities of the words differed by as much as 15 decibels, "but subjects were not usually aware of these differences." Although a speaker's vocal level varied, the mean level

of air pressure in his lungs, found in other measurements to depend on "the loudness with which the speaker was trying to talk," remained "fairly constant" under instructions to count to twenty with equal loudness. The authors conclude that: "naive listeners, obeying an instruction to consider the loudness of sounds in continuous speech, do not assess the acoustic properties of the sounds but consider, instead, the pressure which would be required below the vocal cords" (p. 21).

The most carefully controlled experiment whose findings are seen as favoring the inference of mediating articulation in the perception of stress was performed by Ladefoged and McKinney (1963). The five words: bee, bar, bay, bore, and boo were uttered by one speaker at various vocal levels under conditions that permitted measurement of the peak sound pressure and peak subglottal air pressure during the voicing of each utterance. These utterances were tape recorded and each was spliced into the recorded carrier phrase: "Compare the words bar and..." Thirty listeners were presented with the 65 sentences thus constructed; they were instructed to consider the (constant) loudness of the word bar in the carrier to have a magnitude of "ten" and to state the relative loudness of the terminal word by assigning it a number in the same proportion to ten as its loudness was to the standard, bar. The geometric means of the magnitude estimates were found to be linearly related to the sound pressure level (in log units), slope 1.2, and to the logarithm of the peak subglottal pressure, slope 1.9. The latter relation follows from the authors' earlier finding that the peak subglottal pressure (in cm. of Hg) is proportionate to the 0.6 power of the peak sound pressure (in microbars). The loudness estimates of the

words may also be related to the work done upon the air in producing the voiced sounds. This work (W) is proportional to the product of the subglottal pressure (P) and the rate of flow of air through the glottis (F). In an earlier experiment of this series, the authors showed that P is proportional to F and, hence, $W \propto P^2$. Taking this relation and that between P and sound pressure (S) stated earlier ($P \propto S^{0.6}$), the authors find: $W \propto S^{1.2}$, or $\log W \propto 1.2 \log S$. Since, in log-log coordinates, the magnitude estimates also were related to sound pressure by a straight line with slope 1.2, it follows that the magnitude estimates of the speech stimuli were linearly related to the work done in producing them. The authors conclude: "It would appear that the perceived loudness of words which are within the normal speech range is largely dependent on...the physiological effort required to produce them" (p. 459).

Lane, Catania, and Stevens (1961) have reported that the speaker's estimates of his own vocal level grow as the 1.1 power of the sound pressure he produced. As Ladefoged and McKinney point out, their finding that $W \propto S^{1.2}$ suggests that the Lane et al. speakers were judging physiological effort. This hypothesis is supported by the finding that the "autophonic scale," $\Psi = k S^{1.1}$, was not changed when the speaker was deprived of auditory feedback by intense masking noise. Ladefoged and McKinney conclude that the loudness of a speech sound, whether generated by the listener himself or by an external source, is perceived in terms of the physiological effort required to produce that sound.

The last study of the perception of speech loudness to be cited in support of a motor theory of speech perception was reported by Lehiste and Peterson (1959). A tape recording was prepared in which

various vowels produced with equal effort were mixed at random with vowels produced with unequal effort. The vowels were arranged in pairs at random and presented at the same sound pressure level to listeners instructed to judge the louder of the two vowels in the pair. "Almost invariably, the listeners identified the vowels that were produced with a greater amount of effort as louder than vowels having greater intrinsic amplitude but produced with normal effort" (p. 431). In the light of these findings, the authors suggest that:

"The listener interprets speech according to the properties of the speech production mechanism rather than according to the psychophysical principles of the perception of abstract sounds" (p. 428).

The following kinds of evidence, viewed by the experimenters as favoring the motor theory of speech perception, have been cited in this section: When listeners judge the stress loci in a sentence, these judgments correlate most closely with temporal loci of heightened muscular activity in the speaker, suggesting that the listener is guided by covert articulatory cues in making his estimates. When speakers are instructed to count, making the words equally loud, large changes in intensity are produced although the mean level of air pressure in the lungs remains constant, suggesting that the speaker's criteria for equal loudness depends on proprioceptive cues from muscles involved in articulation. When listeners estimated the loudness of a word uttered at various levels their estimates were linearly related to the work done in uttering the variable word; these estimates were also related linearly to the autophonic scale which describes the speaker's perception

of his own vocal level. Inference from these findings suggests that the listener judges the loudness of speech sounds in terms of the effort cues associated with their production. Finally, when pairs of vowels uttered with unequal effort were presented at equal sound pressure levels and listeners judged the louder vowel in each pair, the vowels produced with greater effort were always judged louder, suggesting that the perception of loudness is more closely correlated with changes in articulation than with the acoustic stimulus.

Evidence opposing the motor theory of speech perception

The inference that speech is perceived by reference to articulation was arrived at, in the studies discussed previously, after examining principally three dependent variables: identification, discrimination, and loudness-estimation functions. The properties of these functions that were repeatedly cited by the investigators as favoring a motor theory of speech perception were: the steepness of the identification gradients, the correlated peaks in the discrimination functions, and the correlation of loudness estimates with vocal-effort measurements.

After considering some methodological limitations of the studies cited above, this section of the review presents evidence that identification, discrimination, and identification-latency functions for consonant, vowel, and nonspeech stimuli do not differ systematically; also, some findings are presented that dispute the presumed relationship between speech loudness and vocal effort.

(1) Some methodological limitations of previous experiments.

Subjects were selected for participation, in several of the experiments reported above, depending on the steepness of their labelling gradients, measured on a pre-test. For example, in the experiment employing the phoneme contrast /do/ - /to/, "the 11 Ss [from a total of 20] whose judgments provided the sharpest phoneme boundaries were selected for the experiment" (Liberman, Delattre, & Cooper, 1958, p. 382); in the /b/ - /d/ - /g/ experiment, "two of the five Ss in group I were eliminated because they failed to apply phoneme labels consistently" (Liberman et al., 1957, p. 360); and in the /b/ - /p/ experiment, "there were 12 subjects in all, chosen from a group of 31 on the basis of a pre-test. The purpose of the selection was to insure that all subjects ultimately serving in the experiment would have a sharp and clear phoneme boundary" (Liberman et al., 1961, p. 181; *italics mine*). Preselection of subjects whose labelling gradients are sharp or steep, according to a criterion that is neither explicit nor quantitative nor uniform, greatly restricts the generality of inferences based on their data. One wonders about the 50 percent of the subject population that was rejected in these three studies; did those Ss perceive speech in a fundamentally different way? The procedure of pre-selecting Ss vitiates, moreover, comparisons of the steepness of labelling gradients collected in different experiments. As a result, it has not been possible to assess the accuracy of an important prediction that follows from the motor theory of speech perception, discussed below; namely, that perception of the consonants is more categorical, i.e., is characterized by steeper labelling gradients, than is perception of the vowels.

The procedure of averaging individual identification and discrimination functions and presenting group data, employed in all the synthetic-

speech studies cited earlier but two, may be placed along side pre-selection of subjects in vitiating comparisons of the steepness of labelling gradients. Since it is generally true that the boundary of the labelling gradient varies, from one S to the next, in its locus along the stimulus continuum, the average gradient is considerably less steep than the individual gradients and reflects none of them well. Put in other terms, this procedure confounds the within-subjects variance, which relates to the judgment of steepness, with the between-subjects variance, which does not.

The judgment of the steepness of an identification gradient, which is crucial to the conclusion of "categorical perception" and the inference of mediating articulation, encounters difficulty even when individual gradients for randomly-selected normal adults are considered. When two gradients, both functions of the same synthetic speech continuum, are compared, the judgment of categorical perception may be based on the extent of the plateaus in response probability or, equivalently, the steepness of the boundary between these plateaus. However, the steepness of a gradient is of interest largely in relation to the steepness of other gradients collected with other stimulus continua: if the steepness of consonant-labelling gradients is cited as evidence of discretely-different mediating responses, then it must be shown that these gradients are more steep than those obtained in comparable experiments with vowels, whose articulations are topographically continuous, and certainly more steep than those obtained with nonspeech stimuli.

When comparing labelling gradients for different stimulus continua, one gradient may not be called more categorical than the other solely because it has more extended regions over which response probability is zero or unity; the extent of these plateaus in the labelling gradient

is largely a matter of experimental convenience. For example, in the experiment whose outcome is shown in Fig. 4, the silent interval for the rabid-rapid continuum could have been allowed to vary only over the range 40 to 90 msec. The outer boundaries of the phoneme classes have not been defined experimentally, so it is not possible to say to what extent the gradients "partition" the total perceptual range.

It appears that the sole property of diverse gradients that will reliably serve as a basis for judgments of steepness is the value of their slope over the range in which it is not zero. However, the slopes of the diverse gradients (or the entire functions for that matter) are obviously not comparable -- despite the widespread practice of comparing them -- since their abscissae are not commensurate. In order to make these comparisons, some normalizing procedure is required; the one employed for this article involves, first of all, determining the distance between each point on the labelling gradient and a step-function on the same coordinates, as illustrated for consonant- and vowel-labelling gradients at the left of Fig. 9. Each positive distance is then divided by the number of units on the stimulus scale marked off by the range between 50 and 75 percent identification; each negative distance is divided by the number of units on the abscissa between 50 and 25 percent identification;⁸ the two functions then may be represented in the same coordinates, as shown at the right of Fig. 9.

Insert Fig. 9 about here

(2) Identification functions for consonants, vowels, and nonspeech stimuli.

The available evidence indicates that identification gradients for consonant stimuli are not reliably steeper than those for vowel stimuli. One such comparison is presented in Fig. 9; although subject to the methodological limitations discussed above, the figure reveals that the two kinds of gradients are quite similar. Clearly-defined boundaries in vowel identification have also been reported by Studdert-Kennedy, Liberman and Stevens (1963), Stevens, Ohman and Liberman (1963), and Bastian and Abramson (1962). These findings conflict with those predicted by the motor theory of speech perception:

"When...acoustic cues are produced by movements that vary continuously from one articulatory position to another (as, for example, the frequency positions of first and second formants produced by various vowel articulations), perception tends to change continuously..." (Liberman et al., 1961, p. 177).

The motor theory maintains that the perception of most consonants is discontinuous or categorical because they covertly evoke discontinuous articulations whose discretely-different sensory feedback controls the overt identifications. On the other hand, the perception of vowels is continuous because the mediating articulations are continuous and provide similar sensory returns. The identification data do not provide a basis for this distinction.

There is considerable evidence to support the conclusions that plateaus in identification are not unique to speech stimuli, but are obtained with diverse continua, and that the identification functions for speech stimuli are not reliably steeper than those for nonspeech stimuli.

In order to compare directly the degree of categorical perception shown by a consonant-labelling gradient with that shown by the generalization gradient for an appropriate nonspeech continuum, Lane and Schneider (1963) conditioned discriminations among the /do/ - /to/ control stimuli employed by Liberman et al. (1961a). It will be recalled that these stimuli were obtained by inverting the /do/ - /to/ spectrograms before converting them to sound, that they were not perceived as speech stimuli, and that their discrimination functions showed only chance levels of accuracy. The two extreme stimuli from this series (nos. 0 and 6) were presented in random order to each of eight listeners serving individually. The subject was seated in front of a tape recorder and counter and was visually isolated from the experimenter. He was told to say /do/ or /to/ in each four-second interstimulus interval, that he would receive a point on the counter for each correct identification, and that each point was worth a penny at the end of the experiment. The experimenter, guided by a protocol, presented the point immediately after each correct response. Discrimination training was terminated after either 30 correct discriminations in succession or 400 stimulus presentations. Three of the eight Ss who participated met the former criterion and were presented individually with the generalization test series, in which the seven control stimuli occurred in an irregular order, with the constraint that each stimulus followed every other exactly twice. In the second through fourth daily experimental sessions, discrimination training was reduced to 100 stimulus presentations and generalization testing was replicated.

The average generalization gradient for the non-speech stimuli, obtained from each of the three Ss, is shown in Fig. 10. These gradients

are directly comparable to the /do/ - /to/ labelling gradients reported

Insert Fig. 10 about here

for 13 Ss by Liberman et al. (1961a). This comparison reveals that the steepness of nonspeech gradients may approach or exceed that for speech gradients. One way of combining the individual gradients in each of these two studies, which is not sensitive to differences in the loci of the gradients on the stimulus continuum, is shown in Fig. 11.

Insert Fig. 11 about here

To illustrate how the functions shown there were obtained, the procedure employed with the individual gradients in Fig. 10 will be described. For each S, the relative frequency with which the /do/ response was emitted to each of the seven nonspeech stimuli was noted as a percentage. Each stimulus was compared to every other in the percent of /do/ responses it evoked, and the difference in percent was obtained for each pair. A frequency distribution of these percent differences was prepared, with class intervals of ten percent, and the frequency distributions for the three Ss were combined. The frequency measures were then converted to relative frequency and plotted in Fig. 11. If the individual gradients were step-functions, the limiting case of "categorical perception", then all the percent differences would be either zero or 100 and the frequency distributions would have peaks in the extreme classes leaving intermediate classes empty.

. Comparison of the frequency distributions in Fig. 11 for identification of: (a) synthetic speech stimuli (Liberman et al., 1961a), (b) groups of tones differing in relative onset times (Lane and

Schneider, 1963), and (c) pure-tone intensities (Cross and Lane, 1962) yields no evidence for the claim that the perception of speech stimuli is peculiarly categorical.

The same conclusion was reached in an experiment comparing identification of visual forms to that of speech stimuli (Cross, Lane, and Sheppard, 1963). Employing circles from which sectors of varying size had been deleted, these investigators conditioned, with one group of Ss, a consonant contrast, /bAb/ - /gAg/, and, with a second group, a vowel contrast, /bAb/ - /bib/, to the extreme stimuli in the series, respectively. In generalization testing, stimuli intermediate in the series were presented and the frequency with which they evoked either identification response was noted. In comparing their findings with nonlinguistic forms to those for the synthetic speech continuum /do/ - /to/ (Figs. 9, 11), the authors reported:

"The identification functions obtained in this study show that response probability remains constant at zero or unity over wide ranges of stimulus variation...the distribution of differences in percent identification of visual stimuli does not differ appreciably from that of speech stimuli, although the set of speech identification gradients selected for comparison is the most 'categorical' of those reported in the literature" (p. 14).

It is also noteworthy that Cross, Lane, and Sheppard found no systematic difference in the identification functions for consonant-contrast responses and vowel-contrast responses.

The findings reported by Lane and Schneider, and Cross, Lane and Sheppard, demonstrate that stimulus control of phonemic contrasts may

be transferred to a nonspeech continuum and that the properties of stimulus identification formerly attributed exclusively to speech perception are preserved by this transfer. If we conclude, therefore, that less than an hour of discrimination training sufficed to transfer the mediation process of speech perception to a nonspeech continuum, then we must ask why the consonant contrasts did not yield more categorical gradients of identification than the vowel contrasts. If it is maintained, on the other hand, that the mediation hypothesis is applicable only to perception along speech continua and is predicated on extensive verbal conditioning (as the motor theorists would have it⁹), then we must ask why the identification functions for nonspeech stimuli so closely resemble those for speech signals.

Since non-linguistic auditory and visual patterns, on the one hand, and synthetic speech stimuli, on the other, both effect "categorical perception," and since mediating articulation is a gratuitous assumption in the former case, it seems unwarranted in the latter.

(3) Discrimination functions for consonants, vowels, and nonspeech stimuli.

The relations between discrimination functions obtained for synthetic speech continua, those predicted from the corresponding identification gradients, and those obtained with comparable nonspeech control stimuli, force an extra degree of complexity on the motor theory of speech perception, which must introduce a direct route for speech discriminations in order to account for the relations among these functions. Concerning the first relationship, illustrated in Fig. 6, the obtained discrimination functions lie above those predicted from the identification gradients. This finding has been interpreted as showing that "listeners are able

in discriminating these stimuli to extract some information in addition to that which is revealed by the way in which they label the stimuli as phonemes" (Liberman et al., 1961, p. 185). This interpretation implies that the listener's speech discriminations are controlled by mediating articulations and by the acoustic stimulus, either alternately or concurrently, and in some relative degree.

Concerning the second relationship, also illustrated in Fig. 6, the speech-discrimination functions lie above those for the nonspeech control stimuli, which are purportedly comparable, both at the phoneme boundary and within phoneme categories. If the listener were relying exclusively on articulatory cues in making the speech discriminations, and if the articulations were truly categorical, then the speech-discrimination functions should fall, within phoneme classes, at the level of or below those functions for the control stimuli. That they do not again forces the interpretation -- if we are to preserve the mediation hypothesis -- that the listener sometimes utilizes proprioceptive information, and sometimes does not. It is unclear by what means the listener manages to use this information just when it has the most to offer, at the phoneme boundary, and to ignore this information when it would detract from discrimination, within phoneme classes.

A consideration of the relations among discrimination functions for consonants, vowels, and nonspeech stimuli provides further evidence that speech discriminations are controlled by the acoustic stimulus directly and not by intervening articulations. Although consonants are generated, in natural language, by discontinuous articulations and vowels are generated by continuous articulations, enhancement of discrimination at the phoneme boundary has been found for both consonant

and vowel stimuli. These peaks in vowel discrimination conflict with the prediction, which follows from the motor theory of speech perception, that perception of the vowels "...tends to change continuously and there are no peaks at the phoneme boundaries" (Liberman et al., 1961, p. 177).

Since the vowel-labelling gradient shown in Fig. 9 is as steep as the consonant-labelling gradient also shown, and since peaks in discrimination were predicted and obtained in the case of the consonant contrast, comparable peaks in vowel discrimination may also be expected. In the article by Fry et al., (1962), a peak in vowel discrimination is indeed predicted in the region of the /*ɛ*/ - /*æ*/ boundary for the two-step and three-step comparisons. Unfortunately, this peak could not be verified by the obtained discrimination data since discrimination over the entire continuum was nearly 100 percent correct in these comparisons. There is, however, some indication of a peak in discrimination in the region of the boundary for the one-step comparisons. In a later experiment employing synthetic vowel stimuli, Liberman, Stevens, and Ohman (1963) have found, on an identification test, "reasonably sharp and well-defined boundaries between phoneme categories," and, on a discrimination test, "peaks at the phoneme boundaries, i.e., discrimination between adjacent stimuli on the vowel continuum is better when the stimuli lie near phoneme boundaries" (p. 13). Bastian and Abramson (1962) did not find peaks, however, in the discrimination of Thai vowels, minimally distinguished by duration. Whether this finding, which contrasts with those just reported and with the data for nonspeech stimuli that follows, should be attributed to some unique property of duration as a discriminative variable, or to a procedural artifact such as near-perfect discrimination over the entire continuum, cannot be determined from the abstract of the paper they presented.

Enhancement of discrimination at the boundaries of identification gradients has been found for nonspeech stimuli as well as for speech stimuli. To return first to the experiment on discrimination of rabid - rapid (Liberman et al., 1961) described earlier, it will be remembered that a second peak was observed in the discrimination function, and that this peak was subsequently shown to correspond to a third "phoneme category," labelled */p/ ("the unnatural /p/"). The obtained peak in discrimination at the /p/ - */p/ boundary is considered, according to the motor theory of speech perception, to reflect differentiated proprioceptive feedback from diverse articulatory patterns evoked by /p/ and */p/. It is difficult to imagine, however, what pattern of articulation has become attached to */p/ over long years of language conditioning, especially since the authors report that "this was a strange and unnatural /p/ to American ears." In this case, the hypothesis of mediating articulation does not provide a convincing account of the observed correlation between labelling and discrimination.

Discrimination functions for nonspeech stimuli collected by Lane and Schneider (1963) provide the most convincing evidence that the correlation between discrimination and labelling does not reflect mediating articulation. These investigators conditioned a discrimination between two nonspeech auditory patterns (the /do/ - /to/ control stimuli), and then measured generalization at intermediate stimulus values, according to the procedure described earlier. Prior to discrimination training in the first experimental session, and after that training in all four sessions, each S was presented with the series of control stimuli arranged in triads for ABX discrimination. Before training, the discrimination functions for the nonspeech stimuli showed chance levels of accuracy.

Following training, these discrimination functions showed peaks at the phoneme boundary in the one- two- and three-step comparisons for all Ss. Table 1 shows that the three listeners discriminated more readily

Insert Table 1 about here

between stimuli drawn from opposite sides of their (nonspeech) "phoneme boundary," shown in Fig. 10, than between stimuli drawn from within a "phoneme category." The labelling and ABX functions for the S who provided the most orderly discrimination data are shown in Fig. 12.

Insert Fig. 12 about here

Since the correlation between discrimination and labelling is observed for nonspeech as well as for speech stimuli, the correlation per se does not seem to warrant the inference that mediating articulation is involved in the discrimination of speech stimuli.

- (4) Gradients of identification latency for consonants, vowels, and nonspeech stimuli.

In 1962, Cross and Lane reported a systematic relationship between the relative frequency and the latency of identification responses evoked by stimuli from a synthetic speech continuum, /do/ - /to/: latencies were minimal at stimulus values corresponding to plateaus in identification (the "phoneme classes") and they were maximal at stimulus values corresponding to the boundary of the identification gradient. Studdert-Kennedy, Liberman and Stevens (1963) have reported comparable findings, both for synthetic speech consonants, /b/, /d/,

/g/, and vowels, /i/, /I/, /ɛ/: "For both stops and vowels, voice and button-press reaction times were lower near phoneme centers than in the vicinity of phoneme boundaries" (p. 33).

Gradients of identification latency for consonant and vowel stimuli are comparable not only to each other but also to latency functions for nonspeech stimuli. Cross and Lane (1962) reported gradients of identification latency for pure-tone intensities that also reached minima within identification plateaus and maxima at the boundary of the identification functions. Lane and Kopp (1963) later replicated these findings. A psychophysical study by Kellogg (1931) obtained a similar relation between identification frequency and latency. Kellogg used seven fixed pairs of visual intensities that varied in the luminance ratio between the members of each pair. His subjects responded either "left side darker" or "right side darker" to each pair and latencies were recorded. When the probabilities of the two responses approached equality, latencies were maximal; when response probabilities were most disparate, latencies were minimal.

In the experiment on identification of visual forms cited earlier, Cross, Lane and Sheppard (1963) also determined the latencies of their subjects, whose responses constituted either a consonant contrast, /bAb/ - /gAg/, or a vowel contrast, /bAb/ - /bib/. These investigators conclude:

"The response-latency functions obtained in this study, with both vowel-contrast and consonant-contrast labelling responses, reach a maximum at the boundary of their corresponding identification gradients and a minimum within the plateaus of those gradients. These findings for response

latencies to non-linguistic visual forms also accord with those reported for identification of synthetic speech stimuli" (p. 15).

The same relationship between identification frequency and latency for nonspeech stimuli was obtained by Cross and Lane (1962), under conditions that permitted measurement of changes in response topography. Two vocal responses whose articulations were continuous were employed: humming with a fundamental frequency of 147 cps and humming with a fundamental frequency of 227 cps. These responses (± 2 cps) were first differentiated out of the subjects' vocal repertoires by selective reinforcement, and then conditioned discriminatively to noise bursts at 50 and 80 db (SPL), respectively. Measurements of response probability, latency, and topography (fundamental frequency), made during subsequent generalization testing, are represented for three subjects in Fig. 13. Inspection of the figure reveals, first of all, the characteristic

Insert Fig. 13 about here

relations between identification frequency and latency obtained in studies employing consonant, vowel, and nonspeech continua. Most important, it will be seen that there was no tendency for the responses to change topographically with changes in the stimulus. This finding with nonspeech stimuli accords with the outcome of an experiment, discussed earlier in this article, by Harris, Bastian, and Liberman (1961). These investigators made electromyographic recordings during mimicry of stimuli from the synthetic speech continuum slit - split. They found that "either there was a normal burst of muscle potential at the lip (indicating a /p/ gesture) or there was not." Since pairs

of vocal responses that are topographically continuous, as well as pairs that are not, fail to show blending at stimulus identification boundaries, it seems that the discreteness of an identification response is more a matter of response stereotypy than it is of articulatory mechanics.

The general finding that the properties of identification, discrimination, latency, and response topography that are obtained with synthetic consonant stimuli are also replicated with synthetic vowels and nonspeech stimuli proves that these properties are not unique to speech perception and strongly indicates that the inference of mediating articulation in speech perception is unwarranted.

- (5) Psychophysical functions for speech loudness and their relation to vocal effort.

Ladefoged has expanded the motor theory of speech perception propounded by Liberman and his co-workers to include perception of the supra-segmental feature of stress; however, he excludes the perception of vowels, discussed by Joos. Ladefoged writes:

"...Vowel quality, nasality, and pitch are most simply related to an ordering of the acoustic properties of the stimulus, whereas most consonant qualities and features such as stress are ordered more simply in terms of articulatory activity" (1962, p. 406).

The evidence seen as favoring this hypothesis has been cited earlier and some countervailing findings may now be described. It is interesting to note in passing that the perception of stress would not be considered to involve mediating articulation according to the motor theory formulated

by Liberman and his collaborators. Since the articulatory correlates of vocal effort or linguistic stress may be presumed to change continuously, changes in stress -- along with changes in other prosodic features and in the vowels -- would probably be correlated directly with changes in the acoustic stimulus.

It was reported earlier that Lane, Catania and Stevens (1961) found that the speaker's estimates of his own vocal level, his autophonic output, grew as the 1.1 power of the sound pressure produced. Then Ladefoged and McKinney (1963) found that listener's estimates of the loudness of speech sounds grew as the 1.2 power of the sound pressure presented. These authors concluded that their listeners were in fact estimating the effort required to produce the sounds that were presented to them.¹⁰ However, Lane, Catania, and Stevens (1961) also obtained listener estimates of speech sounds, varying in autophonic level and in sound pressure level, and came up with a different finding: listener estimates of the loudness of speech signals grew as the 0.7 power of the sound pressure -- a function quite similar to the sone scale of pure-tone loudness.

Lane (1962) employed magnetic-recording techniques to decouple autophonic levels from their corresponding sound pressure levels; the two were then varied independently and the contribution of each to speech-loudness estimates was assessed in a series of eight experiments. When autophonic level was varied from whispering, at the one extreme, to shouting at the other, the quality (spectrum) of the voice inevitably varied but listener estimates of speech loudness remained invariant. These findings show that estimates of the loudness of speech signals are not under the control of the effort involved in producing those signals but rather under the control of the sound pressure levels of the signals.

Further evidence that estimates of loudness are not estimates of effort is provided by experiments that require subjects to match their autophonic level to sounds of varying loudness and by complementary experiments that require subjects to decrease autophonic level, under increases in sidetone level, to maintain sidetone loudness constant. Both kinds of equal-sensation functions, which turn out to be reciprocal power functions, have an exponent ($/0.5/$) predicted by the ratio of the exponents of the autophonic scale and the loudness scale (Lane, Catania and Stevens, 1961). In other words, not one but two disparate psychophysical functions are required to predict the behavior of the subject who listens to the loudnesses of speech signals and matches or regulates them by dint of his own vocal effort.

The final experiment that was cited previously as providing support for the inference of mediating articulation in the perception of stress was conducted by Lehiste and Peterson (1959). These investigators found, it will be recalled, that when two vowels, generated with unequal effort, were presented to listeners at the same sound pressure level, the vowel generated with the greater effort was perceived as louder. The procedure employed in this experiment does not permit us to assess the relative contributions to the perception of loudness from sound pressure level and from effort cues when both are varying, as in natural language. Furthermore, any systematic difference whatsoever between the vowels in each pair may be misconstrued as determining perceived loudness, if the listener is compelled to make loudness judgments in the absence of the acoustic cue normally associated with these judgments. For example, the vowel in each pair with the higher pitch may have been judged louder simply because this was the only discriminable respect in which one vowel was in any way "greater" than the other.

In order to explore the relation between effort cues and sound pressure in determining judgments of vowel loudness, Lane (1962) recorded the phoneme /a/, produced at very low and at very high levels, and played each of these at various intensities to listeners. The slope of the loudness functions, relating estimates to sound pressure levels, was invariant under wide differences in the levels at which the vowels were originally produced. In a subsequent experiment, Lane (1962) had speakers vocalize the phoneme /a/ at a wide variety of levels over a 40-decibel range and then played these vowels at a constant sound pressure level to listeners who were instructed to give magnitude estimates of vowel loudness. Under these conditions, with sound pressure level constant and effort or autophonic level varying, magnitude estimates did increase slightly as a function of the original sound pressure level; the "loudness" scale had a slope of less than 0.1. These findings indicate that the method of pair comparisons employed by Lehiste and Peterson gave a misleading image of the degree to which loudness judgments can be influenced by secondary properties of the vowels; ratio-scaling techniques reveal a small effect which may not be related to vocal effort.

Summary and Conclusions

Vocal responses may be differentiated in topography and then brought under discriminative control either by the language community or by an experimenter in the laboratory. If two discriminative stimuli, drawn from a physical continuum, control two differentiated vocal responses by an observer, and stimuli intermediate on the continuum are presented to him, certain characteristic relations obtain among the following

dependent variables: stimulus identification, stimulus discrimination, identification latency, and response topography.

Identification of stimuli from a synthetic speech continuum, which ranges between two phonemes, constitutes a special case of this general paradigm in which (1) discrimination training and response differentiation have been conducted by the language community, and (2) the identification response may be formally similar to the acoustic stimulus. Under these conditions, it is tempting to speculate that the vocal responses, which in the language community generate the extreme stimuli on the continuum, are active in the perception of those stimuli when they are generated by a device or another person. This hypothesis, that articulation covertly mediates speech perception, has found some support in inferences drawn from the relations among the dependent variables enumerated above.

On the other hand, the inference that articulation mediates perception is unlikely to be formed in comparable experiments in which (1) discrimination training and response differentiation are accomplished quickly in the laboratory, and (2) the identification response is not formally similar to the discriminative stimulus. The same relations among identification, discrimination, latency and topography are found in both kinds of experiments, those employing speech and those employing nonspeech continua, because basically the same conditioning and testing paradigm is employed.

Since the inference of mediating articulation is unfounded in the latter case, we must conclude that it is unwarranted in the former. This conclusion is strengthened by the finding that the inference of mediating articulation (1) is based on data whose generality is often

seriously constrained by methodological limitations, and (2) leads to certain differential predictions which have not been confirmed by experiment.

Footnotes

1. The preparation of this manuscript and several of the experiments reported herein were supported by a contract with the Language Development Section, United States Office of Education. The assistance of Dr. N. J. Peterson, D. V. Cross, and B. A. Schneider is gratefully acknowledged.

2. In the most recent explication of the theory, Liberman et al. (1963) suggest that the mediating events may be one or two steps removed from articulation: "The neural commands themselves, or rather their equivalents in the central nervous system, might be used to provide the reference system in terms of which the decoding [of speech] is carried out" (p. 8).

3. In linguistic terms, Liberman et al. (1963) have put it this way: "The articulatory reference theory [rests] empirically on the evidence which indicates that the relation between phoneme and articulation is more nearly one-to-one than is the relation between phoneme and acoustic signal" (p. 7).

4. For a description of Pattern Playback and its principles of operation, see Cooper 1950, 1953 and Cooper, Liberman & Borst, 1951.

5. Certain additional modifications were made in the control patterns "to provide a further precaution against the possibility that the control stimuli would resemble speech" (Liberman et al., 1961a, p. 381).

6. "The anomalous status of /k/ is presumably to be explained on the basis of some discontinuity, not at the articulatory level, but at the point of conversion of the articulatory gesture to acoustic output. It would appear that the acoustic description provides a poorer

approximation to the ideal phoneme than does the articulatory characterization, at least for /k,g/" (Lisker, Cooper, and Liberman, 1962, p. 94).

7. Sometimes a vowel quadrilateral is used as a model; this does not change Joos' theory substantially.

8. The stimulus ranges between .5 and 1.0, and between .5 and 0.0 response probability were not employed for normalizing, since a small error of measurement would have a very large effect on the estimates of steepness.

9. "We believe that in the course of his long experience with language a speaker (and listener) learns to connect speech sounds with their appropriate articulations" (Liberman et al., 1961, p. 177).

10. Whereas Ladefoged and McKinney contend that estimates of speech loudness are actually estimates of vocal effort, Warren (1962) has recently argued that estimates of vocal effort are actually estimates of loudness.

References

- Abramson, A. S. Identification and discrimination of phonemic tones. J. Acoust. Soc. Amer., 1961, 33, 842 (abstract).
- Bastian, J. & Abramson, A. S. Identification and discrimination of phonemic vowel duration. J. Acoust. Soc. Amer., 1962, 34, 743-744 (abstract).
- Bastian, J., Eimas, P. D., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. J. Acoust. Soc. Amer., 1961, 33, 842 (abstract).
- Cooper, F. S. Spectrum analysis. J. Acoust. Soc. Amer., 1950, 22, 761-762.
- Cooper, F. S. Some instrumental aids to research on speech. In Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching. Wash., D.C.: Inst. Lang. Ling., Georgetown Univer., 1953.
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. J. Acoust. Soc. Amer., 1952, 24, 597-606.
- Cooper, F. S., Liberman, A. M., & Borst, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proc. Nat. Acad. Sci., 1951, 37, 318-325.
- Cooper, F. S., Liberman, A. M., Harris, K. S. & Grubb, P. M. Some input-output relations observed in experiments in the perception of speech. Proc. Second Intl. Congr. on Cybernetics. Namur, Belgium: 1958.
- Cross, D. V. & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. J. exp. Anal. Behav., 1962, 5, 487-496.

- Cross, D. V., Lane, H. L., & Sheppard, W. C. Gradients of response probability and latency for discrimination and generalization of visual forms. Unpublished report, Behavior Analysis Laboratory, Univ. of Michigan, 1963.
- Draper, M. H., Ladefoged, P. & Whitteridge, D. Respiratory muscles in speech. J. Speech Hear. Res., 1959, 2, 17-27.
- Fant, G. Acoustic Theory of Speech Production. The Hague: Mouton and Co., 1960.
- Fry, D. B., Abramson, A. S., Eimas, P. D. & Liberman, A. M. The identification and discrimination of synthetic vowels. Lang. Sp., 1962, 5, 171-189.
- Harris, Katherine S., Bastian, J. & Liberman, A. M. Mimicry and the perception of a phonemic contrast induced by silent interval: electromyographic and acoustic measures. J. Acoust. Soc. Amer., 1961, 33, 842 (abstract).
- Hirsch, I. J. Auditory perception of temporal order. J. Acoust. Soc. Amer., 1959, 31, 759-767.
- Jones, S. The accentⁱⁿ French -- what is accent? La Maitre Phonetique, 1932, 40, 74-75.
- Joos, M. Acoustic Phonetics. Language, 24, 1948, Monograph No. 23.
- Kellogg, W. N. The time of judgment in psychometric measures. Amer. J. Psychol., 1931, 43, 65-86.
- Ladefoged, P. The Perception of Speech. In: Mechanization of thought Processes. London: H. M. Stationery Office, 1962.
- Ladefoged, P., Draper, M. H., & Whitteridge, D. Syllables and stress. Miscellanea Phonet., 1958, 3, 1-14.

- Ladefoged, P. & McKinney, N. P. Loudness, sound pressure, and subglottal pressure in speech. J. Acoust. Soc. Amer., 1963, 35, 454-460.
- Lane, H. L. Psychophysical parameters of vowel perception. Psychol. Monogr., 1962, 76, 1-25.
- Lane, H. L., Catania, A. C. & Stevens, S. S. Voice level: autophonic scale, perceived loudness, and effects of sidetone. J. Acoust. Soc. Amer., 1961, 33, 160-177.
- Lane, H. L. & Kopp, J. L. The effects of response dependent and independent reinforcement in extending stimulus control. Psychol. Rec., 1963, in press.
- Lane, H. L. & Moore, D. J. Reconditioning a consonant discrimination in an aphasic: an experimental case history. J. Speech Hear. Disord., 1962, 27, 232-243.
- Lane, H. L. & Schneider, B. A. Discriminative control of concurrent responses by the intensity, duration and relative onset time of auditory stimuli. Unpublished report, Behavior Analysis Laboratory, Univ. of Michigan, 1963.
- Lehiste, Ilse & Peterson, G. E. Vowel amplitude and phonemic stress in American English. J. Acoust. Soc. Amer., 1959, 31, 428-435.
- Liberman, A. M. Some results of research on speech perception. J. Acoust. Soc. Amer., 1957, 29, 117-123.
- Liberman, A. M., Cooper, F. S., Harris, K. S. & MacNeilage, P. F. A motor theory of speech perception. Proc. Speech Commun. Seminar, Royal Institute of Technology, Stockholm, 1963.
- Liberman, A. M., Delattre, P. & Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. Amer. J. Psychol., 65, 1952, 497-516.

- Liberman, A. M., Delattre, P. C., & Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. Lang. Speech, 1958, 1, 153-157.
- Liberman, A. M., Delattre, P. C., Cooper, F. S. & Gerstman, L. J. The role of consonant-vowel transitions in the perception of stop and nasal consonants. Psychol. Monogr., 1954, 68, 13 pp. (Whole No. 379).
- Liberman, A. M., Harris, Katherine, S., Eimas, P., Lisker, L., & Bastian, J. An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. Lang. and Sp., 1961, 4, 175-195.
- Liberman, A. M., Harris, Katherine S., Hoffman, H. S. & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol., 1957, 54, 358-368.
- Liberman, A. M., Harris, Katherine S., Kinney, JoAnn, & Lane, H. L. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. J. exp. Psychol., 1961, 61, 379-388. (a)
- Liberman, A. M., Stevens, K. N., Ohman, S. Identification and discrimination of synthetic speech sounds. Unpublished report, STL-QPSR-2, Royal Institute of Technology, Stockholm, 1963.
- Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. Lang., 1957, 33, 42-49.
- Lisker, L., Cooper, F. S. & Liberman, A. M. The uses of experiment in language description. Word, 1962, 18, 82-106.
- Miller, G. A. The magical number seven, plus or minus two. Psychol. Rev., 63, 1956, 81-97.

- Peterson, G. E. & Barney, H. L. Control methods used in a study of the vowels. J. Acoust. Soc. Amer., 1952, 24, 175-184.
- Stevens, K. N., Ohman, S. E. G., & Liberman, A. M. Identification and discrimination of rounded and unrounded vowels. Paper read at the 66th meeting of the Acoustical Society of America. Ann Arbor, 1963.
- Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. Paper read at the 66th meeting of the Acoustical Society of America. Ann Arbor, 1963.
- Warren, R. M. Are "autophonic" judgments based on loudness? Amer. J. Psychol., 1962, 75, 452-456.

Table 1

Average percent correct discrimination of nonspeech stimuli, presented in ABX triads to three Ss.

Comparison	For A and B stimuli drawn from the same "phoneme class"*		For A and B stimuli drawn from different "phoneme classes"*		χ^2	P
	%	N	%	N		
one-step	49.8	239	65.6	63	8.04	< .05
two-step	66.7	192	80.7	155	11.37	< .01
three-step	71.1	91	85.5	219	11.46	< .01

*The difference in percent identification of A and B stimuli was always either less than 25 percent (same class) or greater than 60 percent (different classes).

Figure Captions

Fig. 1. The frequency-positions of the schematic stops (A) and of the vowel formants (B) that were combined (as in C) and converted into sound for identification by listeners. (After Liberman, Delattre and Cooper, 1952).

Fig. 2. The spectrographic patterns, varying in the direction and extent of second-formant transition, that were converted into sound for identification as /b/, /d/, or /g/. (From Liberman et al., 1957).

Fig. 3. Identification of the consonants synthesized with the spectrographic patterns shown in Fig. 2. Each point shows the relative frequency of the responses /b/, /d/, and /g/, in 32 presentations of each stimulus to one listener. (After Liberman et al., 1957).

Fig. 4. Identification of synthetic speech stimuli as rabid or rapid, as a function of the duration of the silent interval between syllables. Each point shows the mean relative frequency of the responses /b/ and /p/ in 32 presentations of each stimulus to 12 listeners. (From Liberman et al., 1961).

Fig. 5. Discrimination of the synthetic speech stimuli shown in Fig. 2, when presented in ABX triads to one listener. The predicted functions are based on the labelling gradients for this S, shown in Fig. 3. Each point on the obtained functions shows the relative frequency of correct discriminations in 42 presentations of each triad. (After Liberman et al., 1957).

Fig. 6. Discrimination of the synthetic speech stimuli rabid-rapid (filled circles), and comparable nonspeech stimuli (crosses), when presented in ABX triads to 12 listeners. The "expected" functions are based on the labelling gradients obtained from these Ss when they

were instructed to identify each stimulus as /b/, /p/, or */p/. Each point on the obtained functions shows the mean relative frequency of correct discrimination in 40 presentations of each triad. (From Liberman et al., 1961).

Fig. 7. The spectrographic patterns that produce /d/ and /g/ before various vowels. The dashed lines are extrapolations to the /d/ and /g/ loci. (From Liberman, 1957).

Fig. 8. The frequency of the second formant versus the frequency of the first formant for ten vowels by 76 speakers. (From Peterson and Barney, 1952).

Fig. 9. Comparison of consonant and vowel identification with synthetic speech stimuli. Top left: mean relative frequency of the /do/ response (/do/ - /to/ contrast) in 6 presentations of each stimulus to 27 Ss. (After Liberman, Delattre, and Cooper, 1958). Bottom left: mean relative frequency of the /æ/ response (/æ/ - /ɛ/ contrast) in a variable number of stimulus presentations (84-184) to eight Ss. (The complement of the gradient reported by Fry et al., 1962, is shown.) Right: Normalized gradients of consonant and vowel identification (for normalizing procedure, see text).

Fig. 10. Identification of nonspeech stimuli. The seven spectrographic patterns in the series employed by Liberman et al. (1961a) to study /do/ - /to/ identification were inverted and then converted to sound. Each of three Ss was conditioned to respond "do" to the stimulus with no cutback of the first formant and "to" to the stimulus with 60 msec cutback. In generalization testing, each of the seven stimuli was presented 60 times, in irregular order.

Fig. 11. Comparison of the degree of "categorical perception" obtained in identification of synthetic speech stimuli (Liberman et al., 1961a), comparable nonspeech stimuli (see Fig. 10), and pure tone intensities (Cross and Lane, 1962). For each experiment, each stimulus was compared to every other in percent identification and the relative frequency of within-pairs differences was plotted. In the limiting case of "categorical perception", the labelling gradients would have the form of a step-function and the frequency distribution would be nonzero only in the extreme classes.

Fig. 12. Discrimination of nonspeech (control) stimuli when presented in ABX triads to one listener. Each point is the relative frequency of correct discrimination in 24 presentations of pairs of stimuli separated by one, two, or three steps in the series of seven stimuli (open circles, closed circles, and hexagons, respectively). This S's identification gradient, which appears in Fig. 10 (S2), is also plotted (triangles) to reveal that peaks in discrimination occur at the "phoneme boundary" of the nonspeech stimuli.

Fig. 13. Relations among response probability, latency, and topography in stimulus generalization. Top: Median frequency in cps (circles) and the range (vertical lines) of high- and low-pitch responses as a function of stimulus intensity for each of three Ss. The dashed horizontal lines represent the vocal pitches previously differentiated. Bottom: Response probabilities (squares) equal the ratio of the number of low-pitch responses emitted to the number of stimulus presentations (10) at each intensity. The hexagons represent the average latency of high- and low-pitch responses to each stimulus intensity.

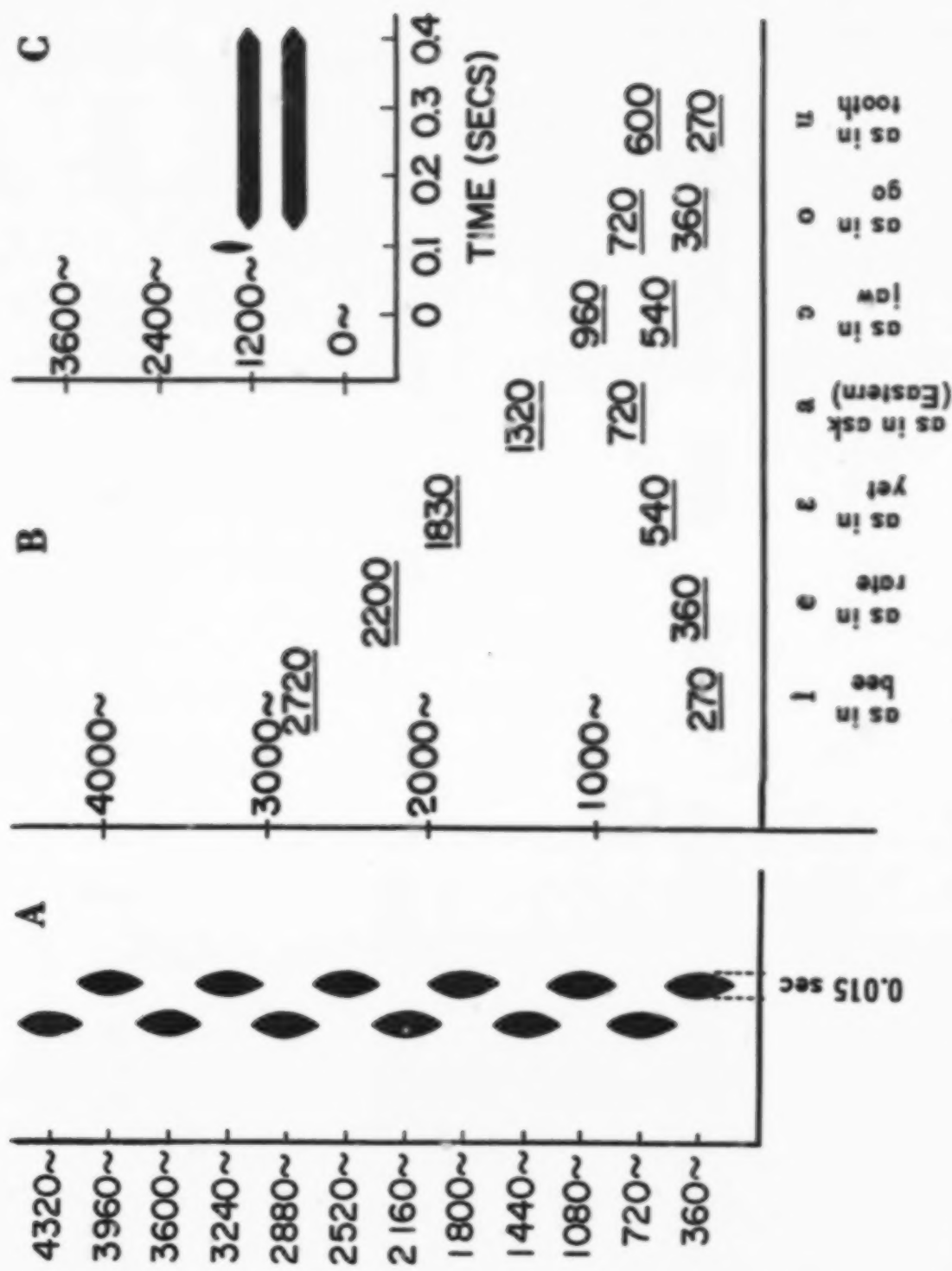
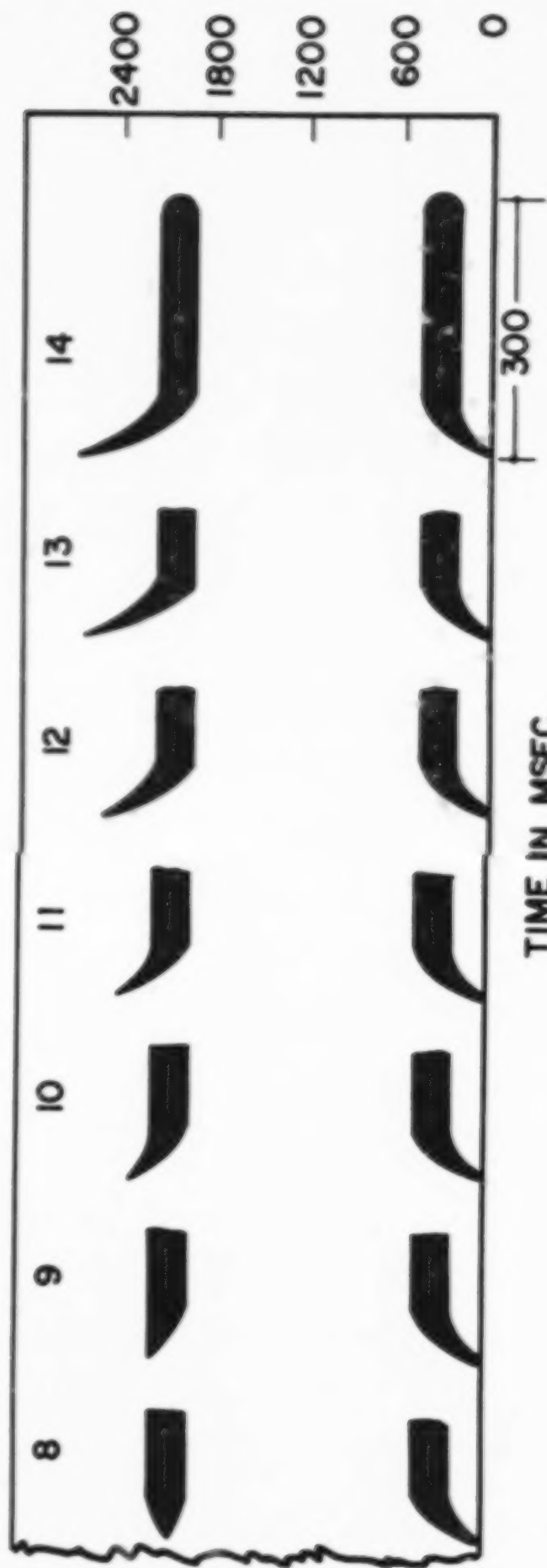
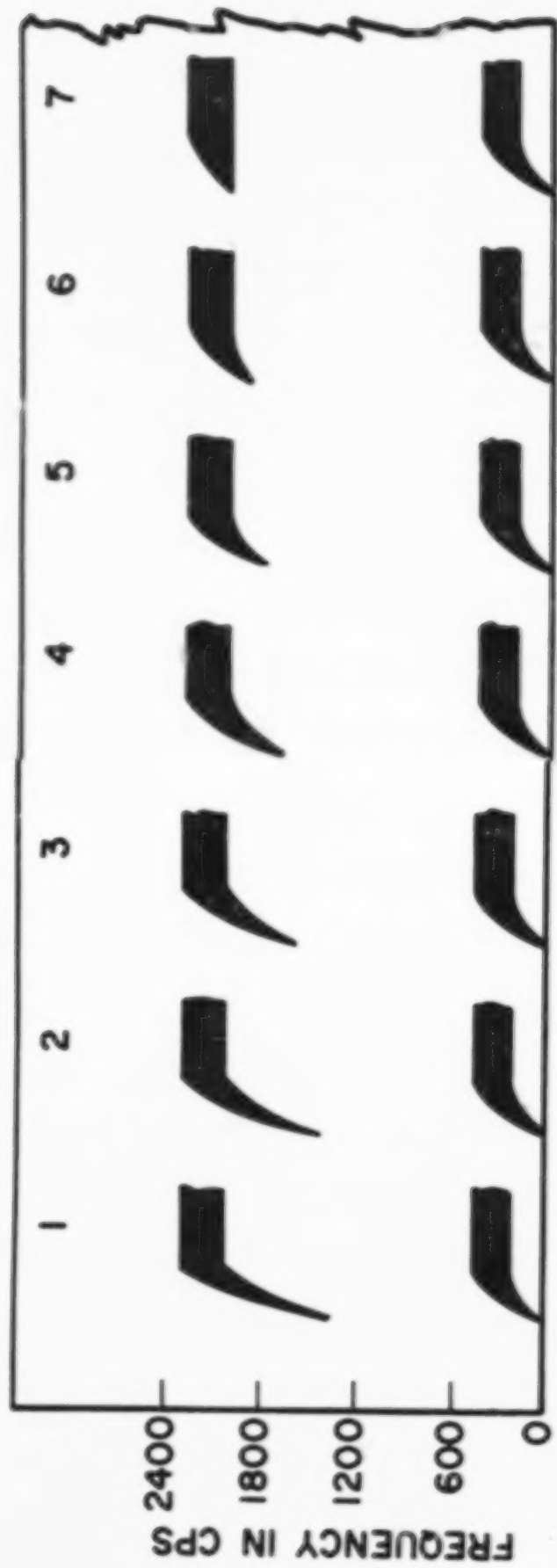


Fig. 1



TIME IN MSEC.

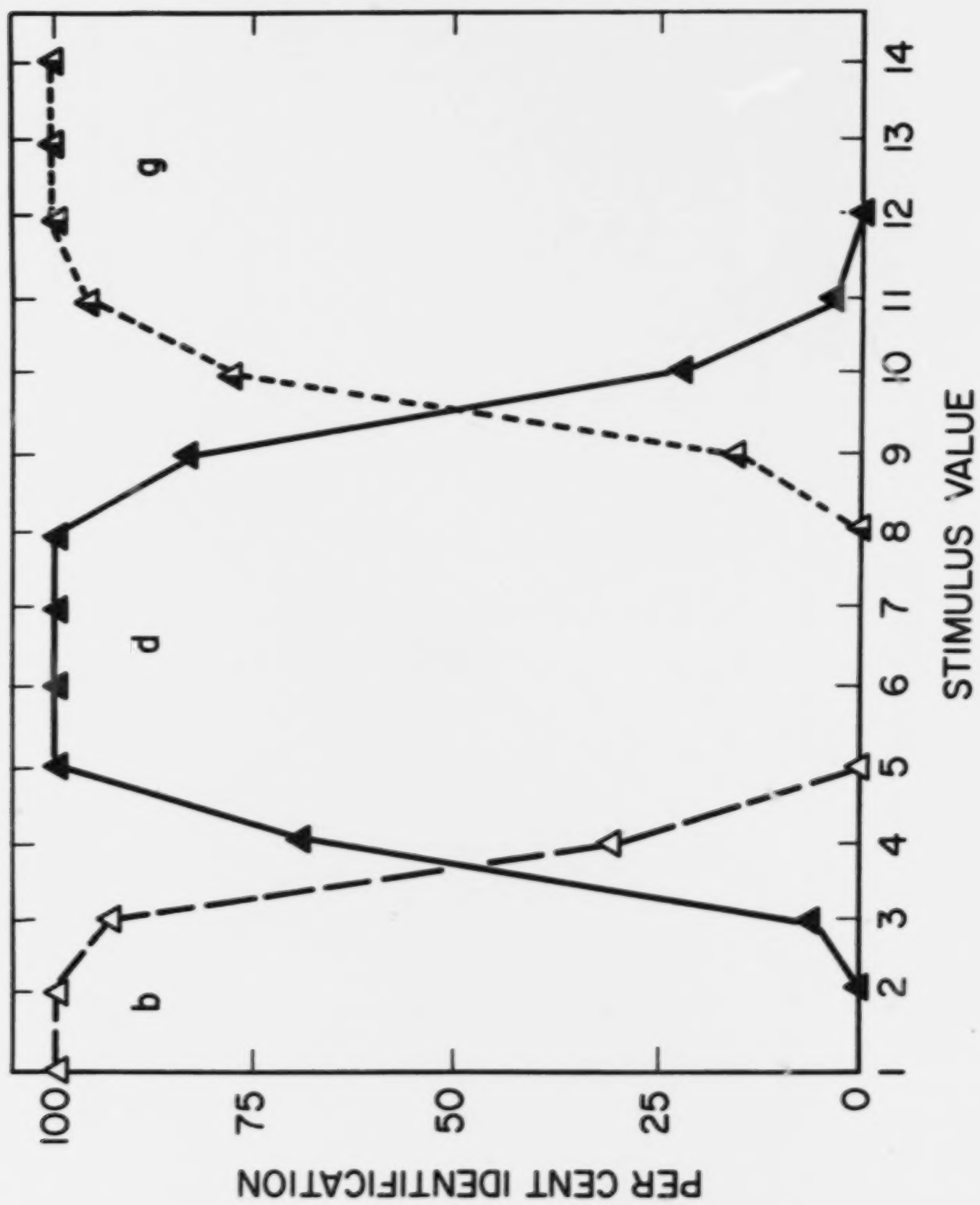


Fig. 3

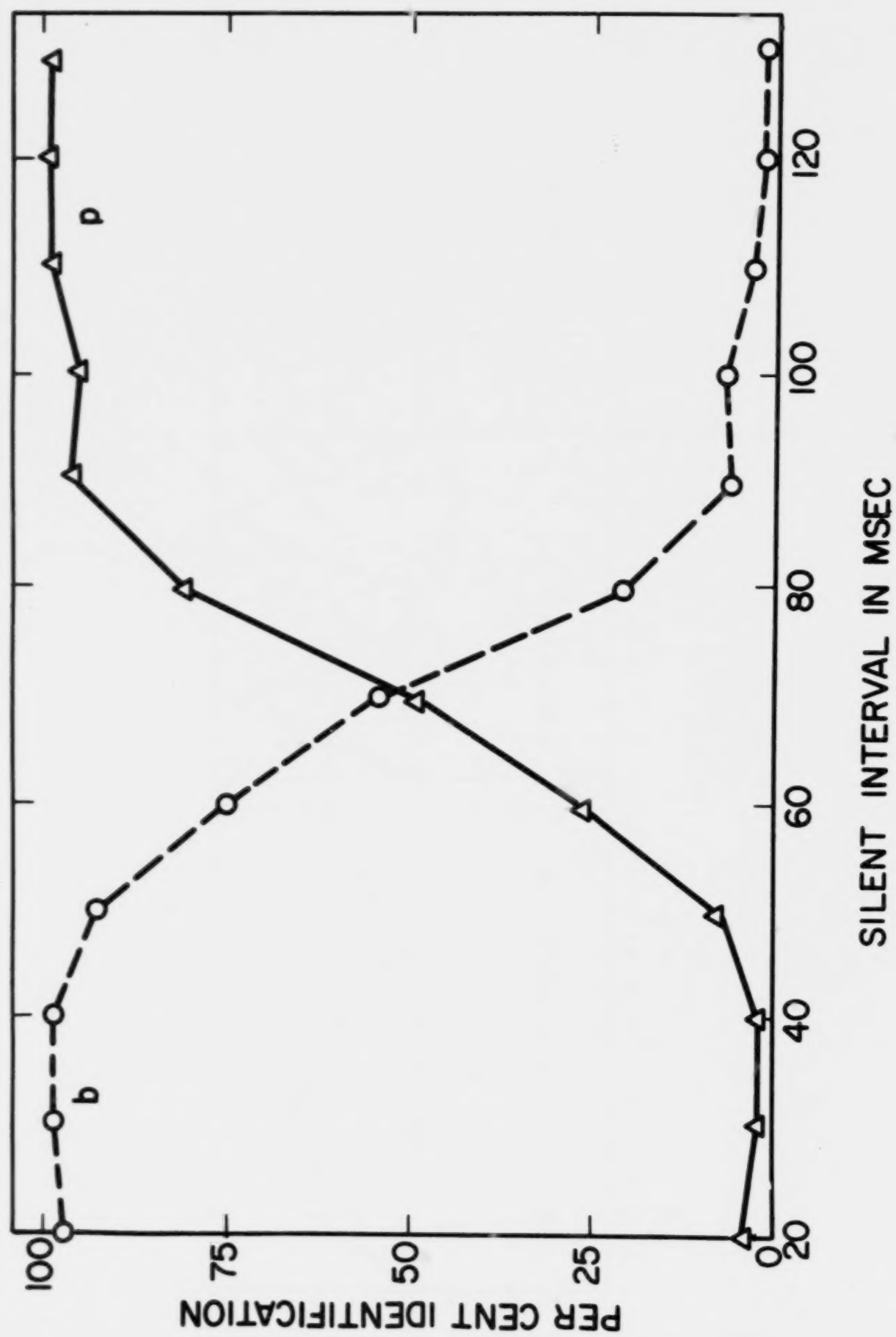
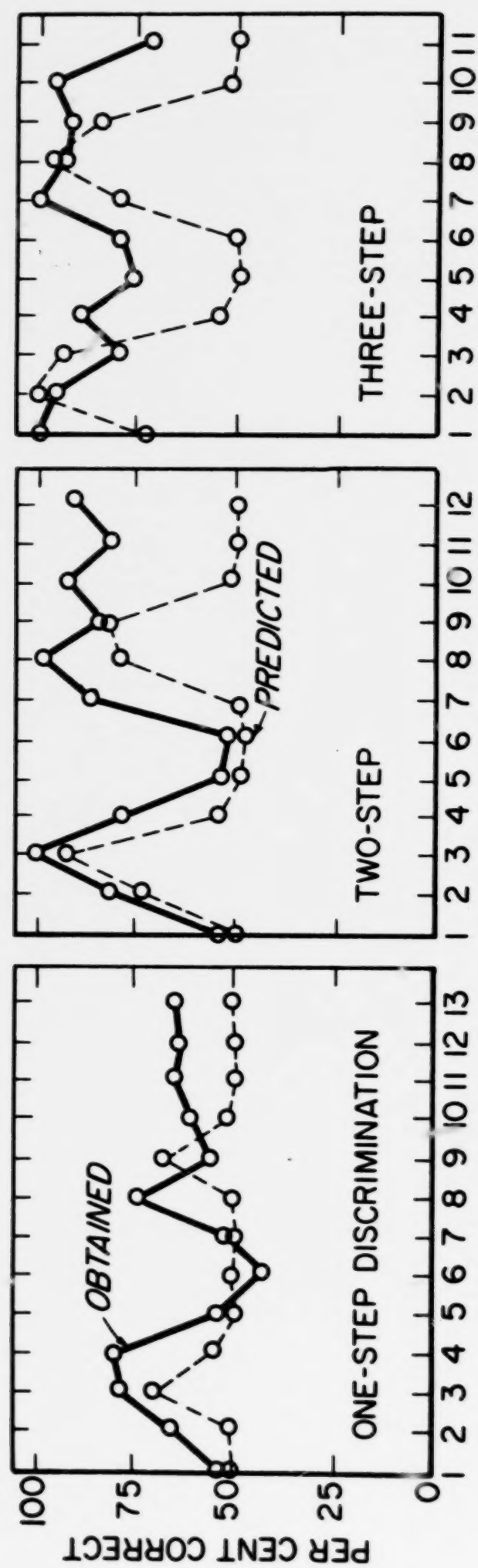


Fig. 4



VALUE OF "A" STIMULUS

Fig. 5

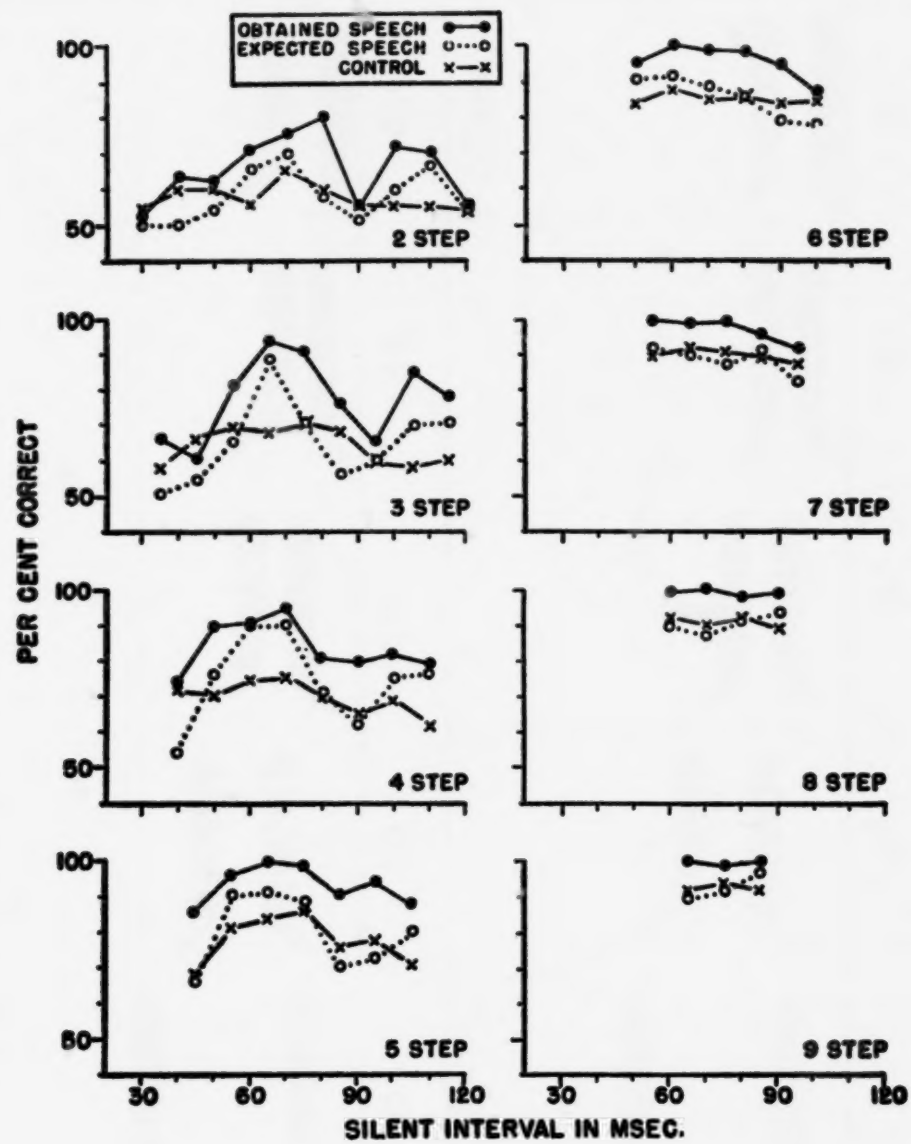


Fig. 6

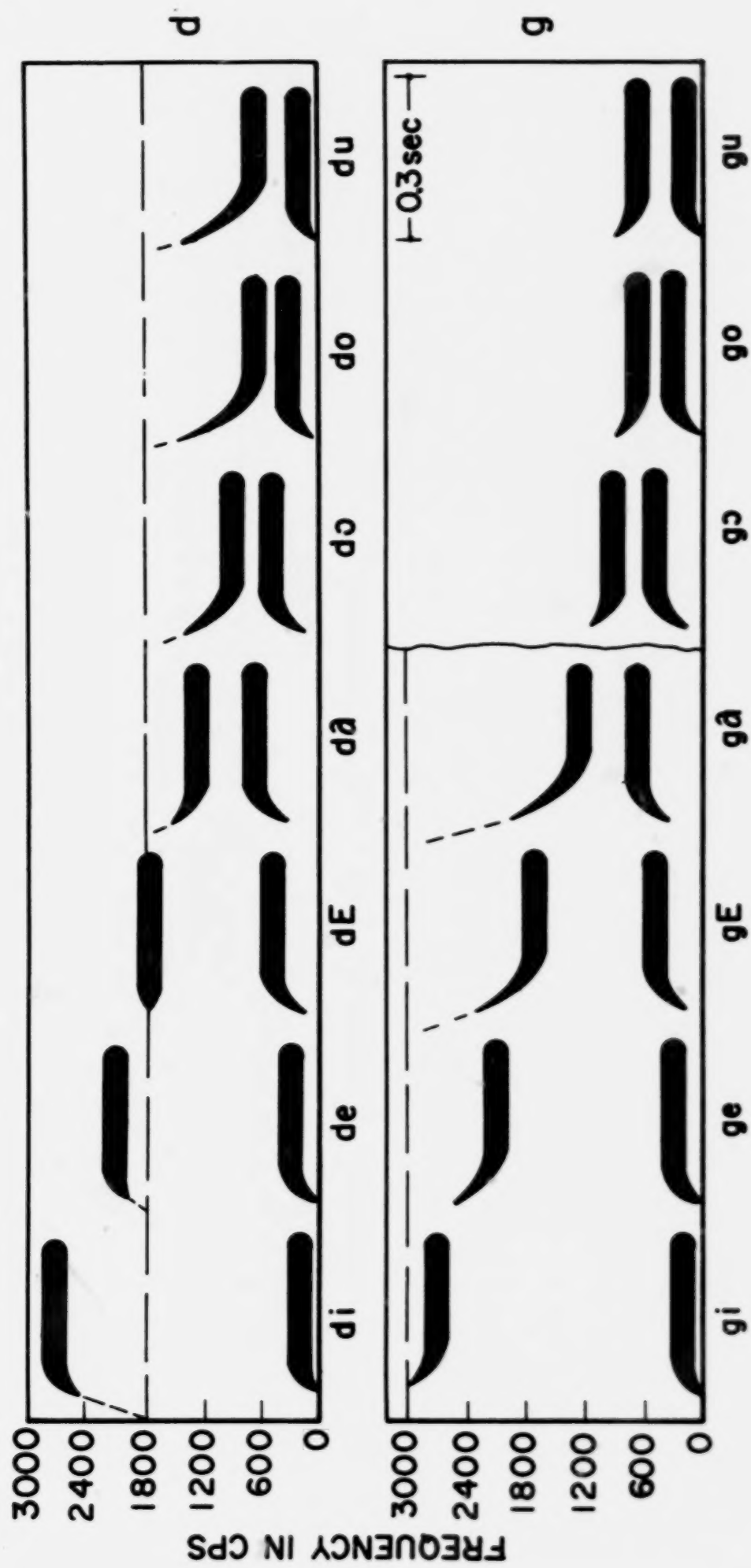


FIG. 7

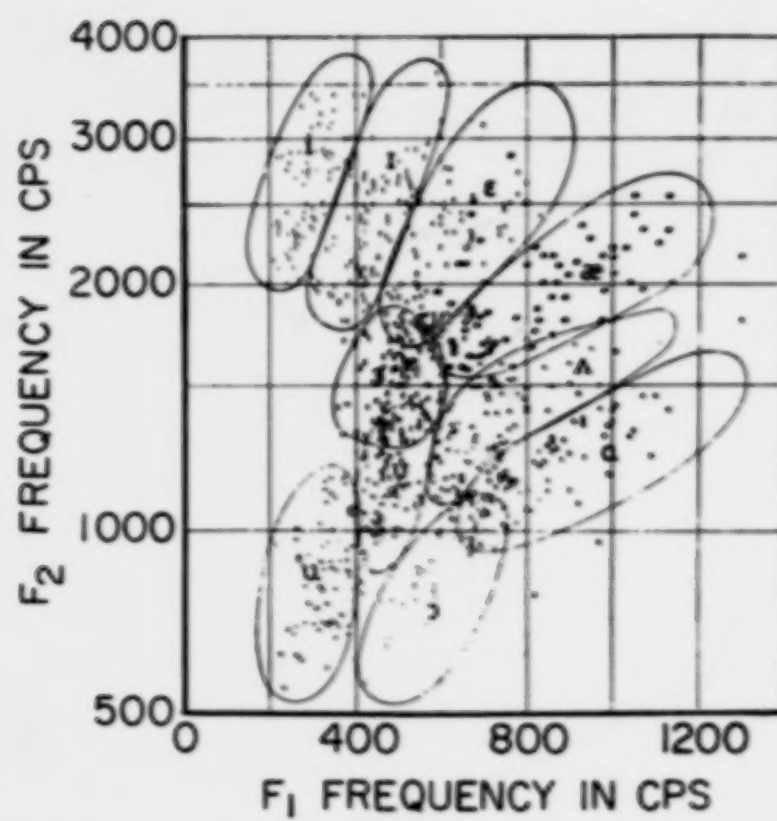


Fig. 8

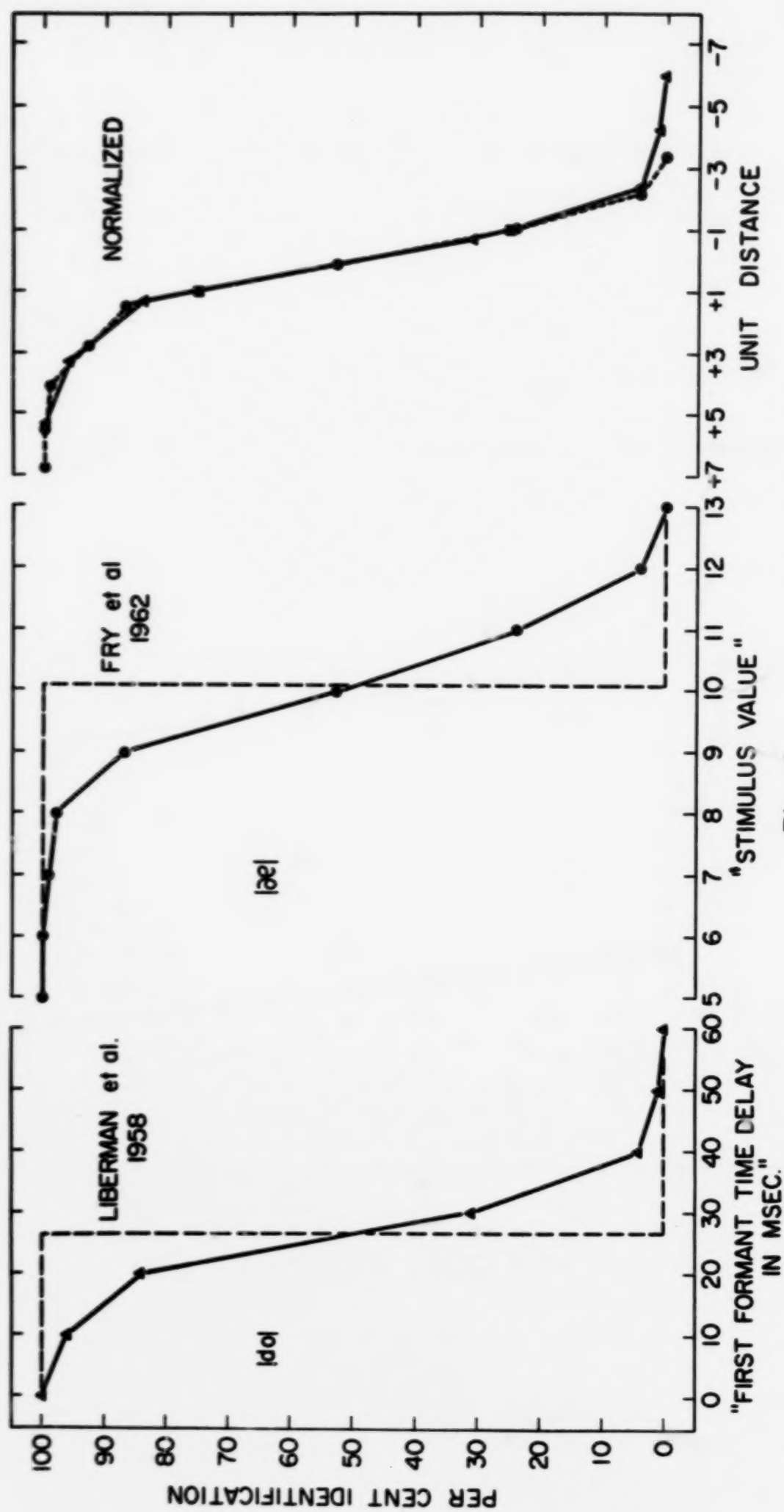


Fig. 9

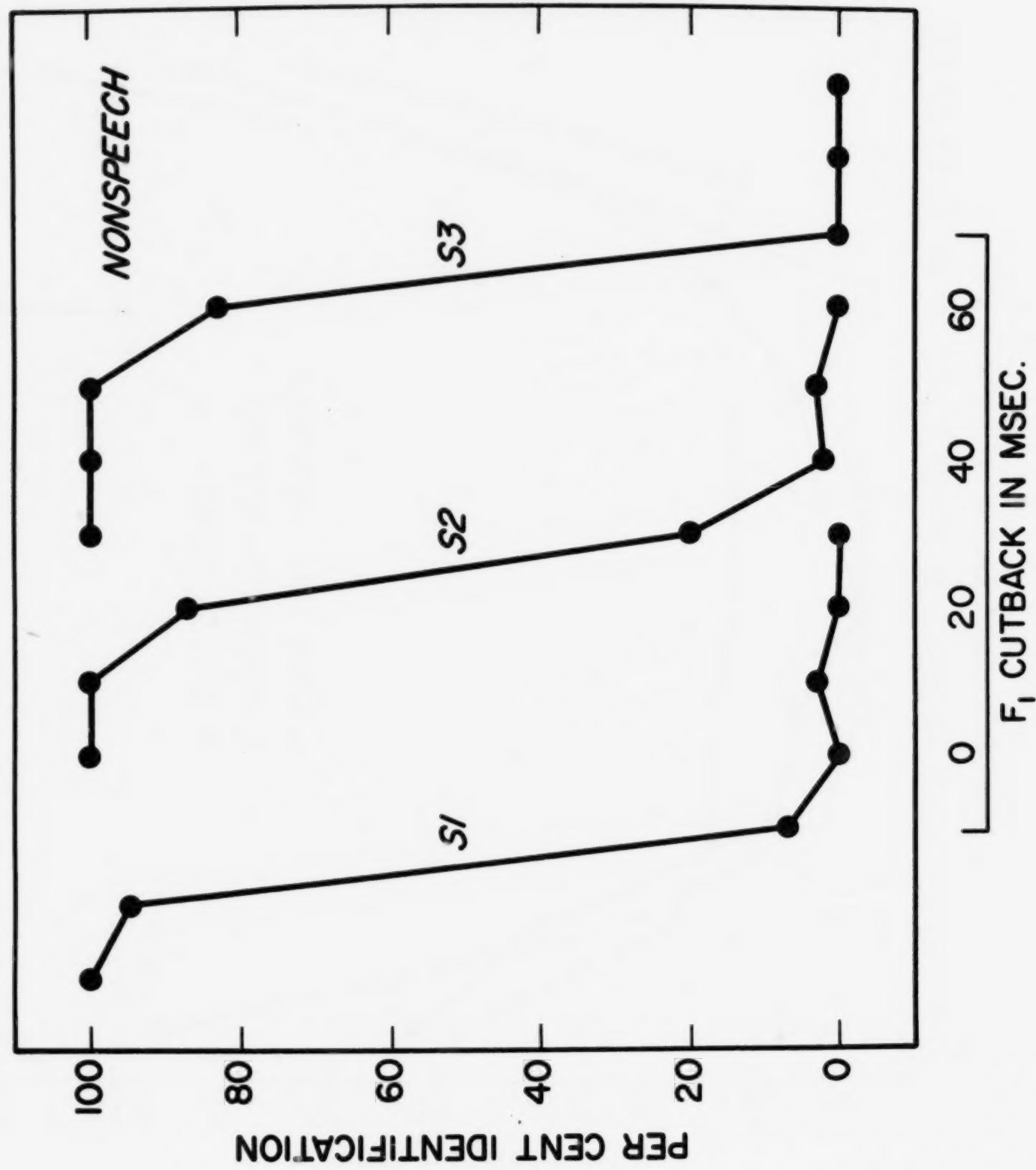


Fig. 10

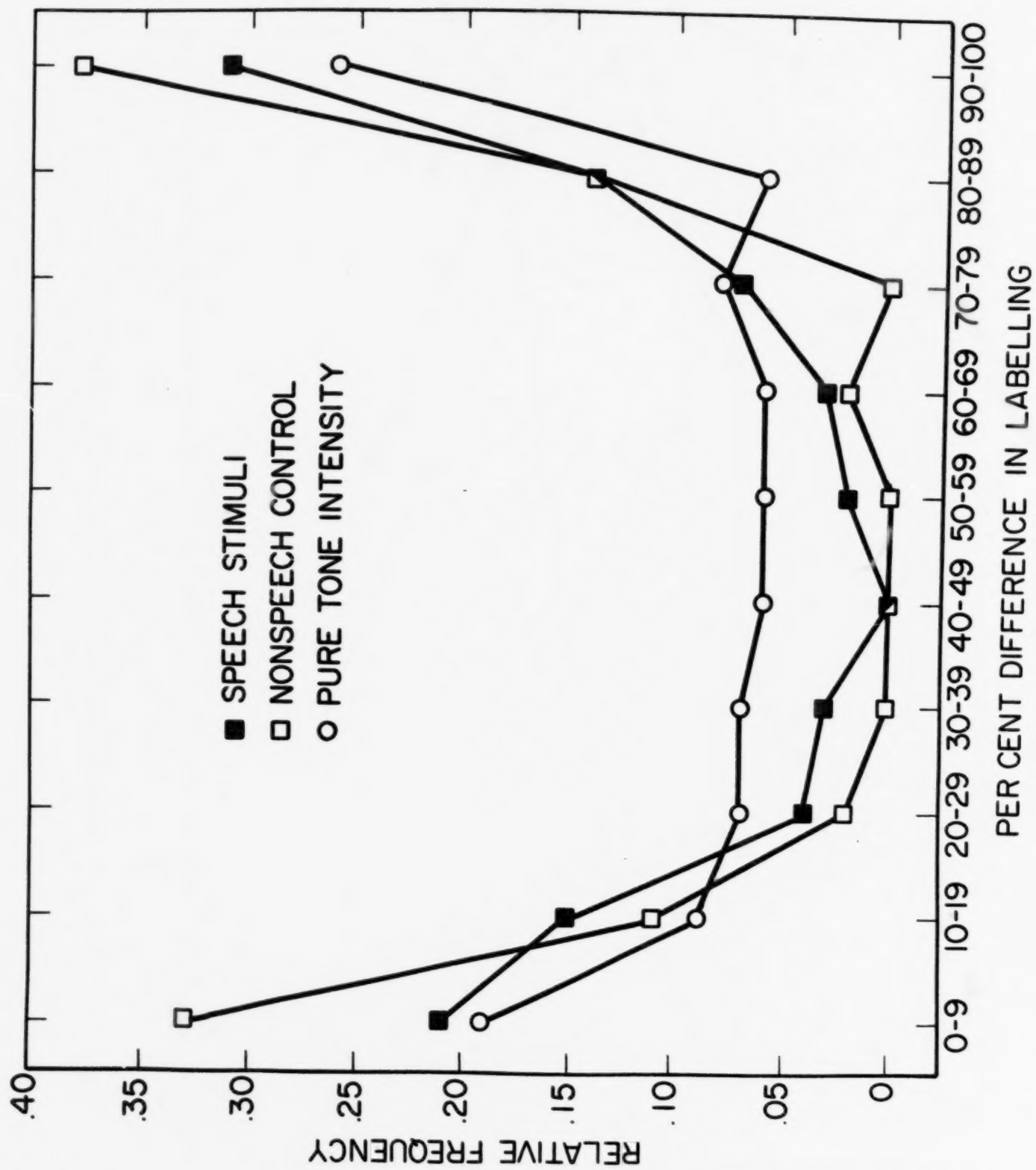


Fig. 11

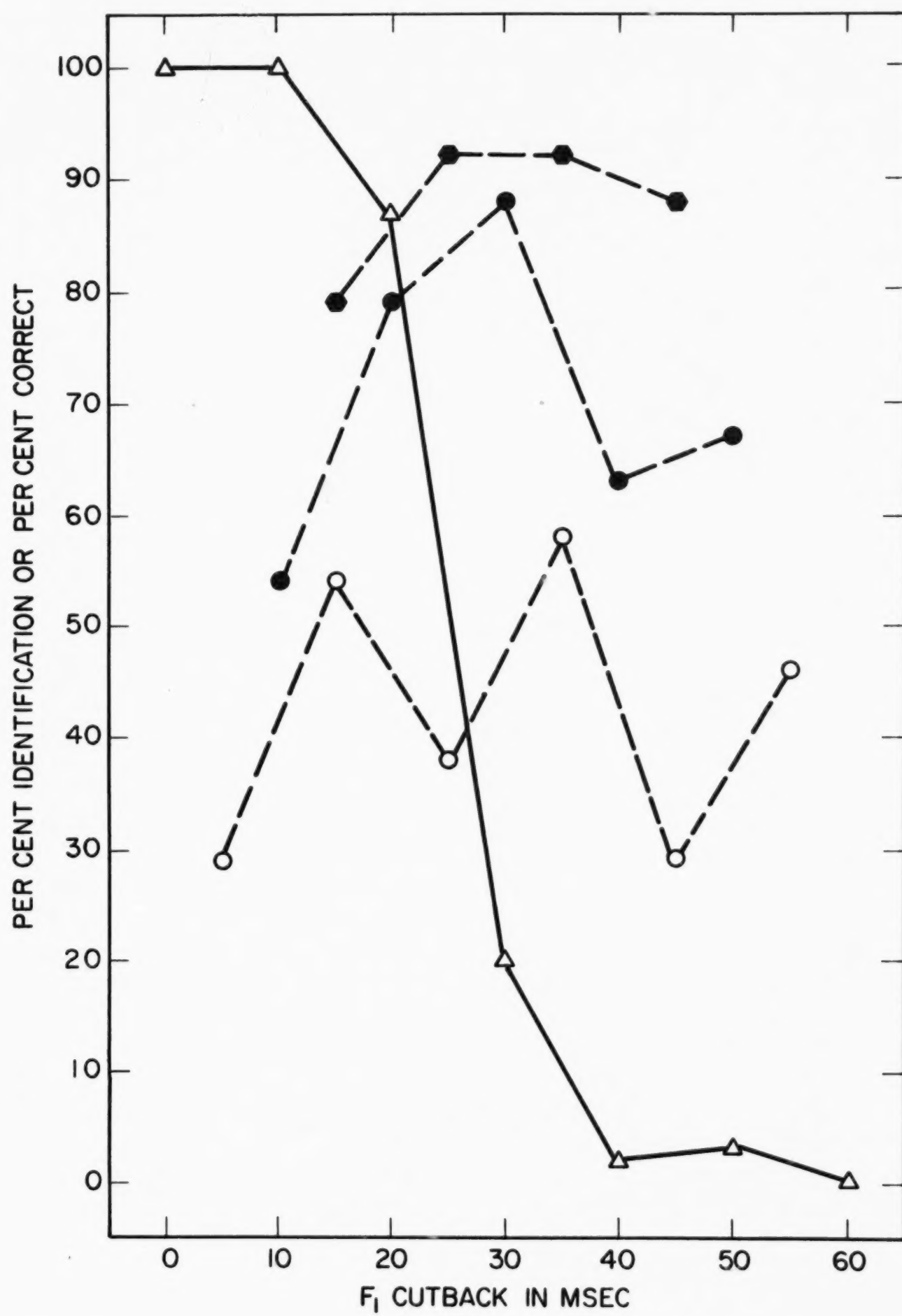


Fig. 12

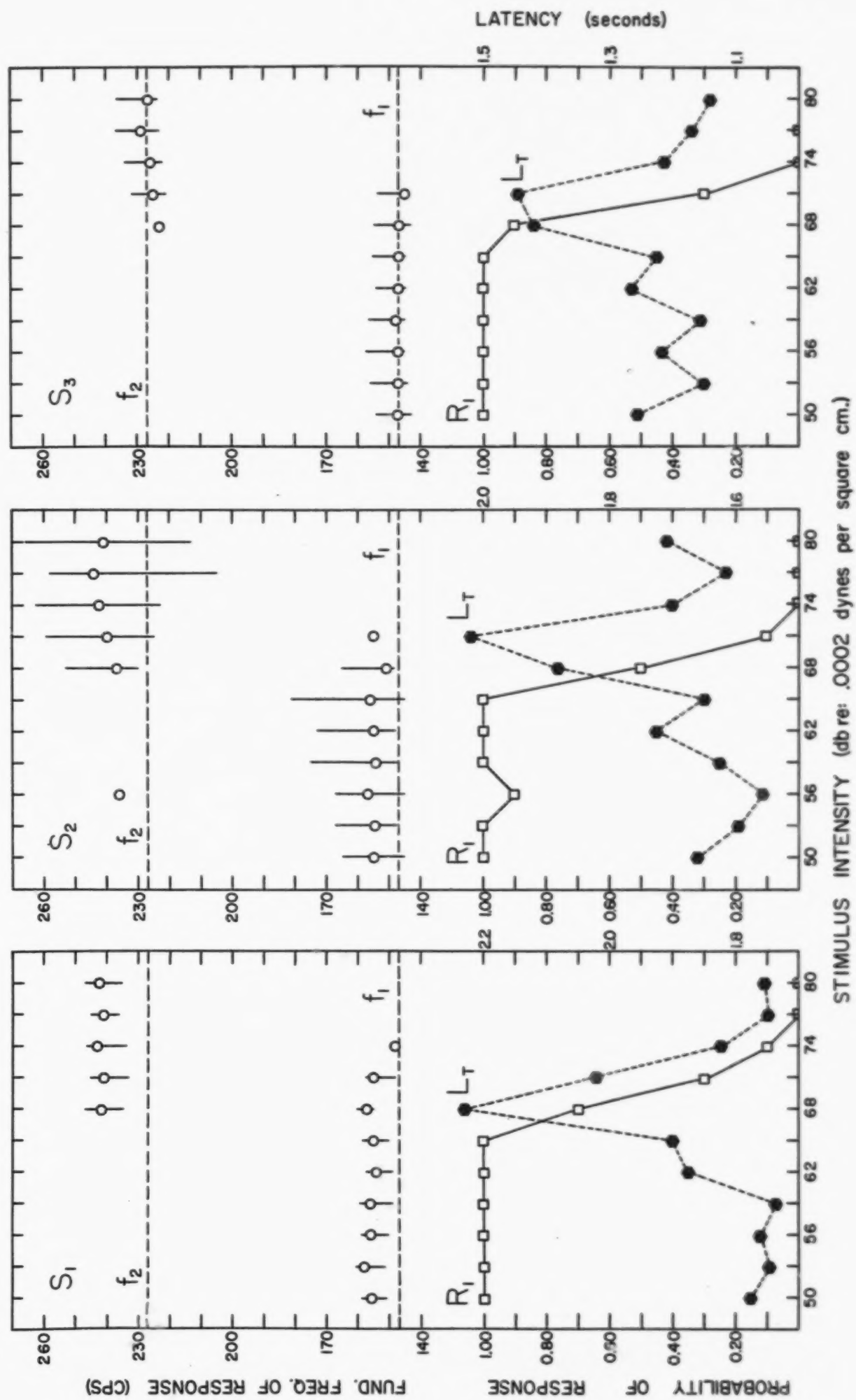


Fig. 13

VISUAL CONTROL OF PHONEMIC CONTRASTS: A TEST OF
THE MOTOR THEORY OF SPEECH PERCEPTION

David V. Cross, Harlan Lane and William Sheppard
The University of Michigan

The Visual Control of Phonemic Contrasts: a Test
of the Motor Theory of Speech Perception¹

D. V. Cross, H. L. Lane, and W. C. Sheppard

Behavior Analysis Laboratory

University of Michigan

Advocates of the motor theory of speech perception contend that "speech sounds are perceived by reference to the articulatory movements that produce them, and this articulatory reference is important for the distinctiveness of speech as perceived" (Liberman et al., 1963). According to this view, the sounds of speech are absolutely and quickly identifiable because continuous variations in the acoustic stimulus cue phonemic distinctions that are categorical in nature for most speech continua. The theory maintains that the categorical perception of speech arises from a learned connection between speech sounds and the articulations that generate them, and that "in time, these articulatory movements come to mediate between the incoming acoustic stimulus and its ultimate perception" (op. cit.). Stimuli from the same phoneme class evoke the same mediating response and acquire similarity, whereas stimuli from different phoneme classes evoke different mediating responses and gain the added distinctiveness of their different proprioceptive returns. Thus, speech perception is distinctive because it is categorical, and it is categorical because the mediating articulations are categorically different.

Some of the evidence adduced in support of this view comes from studies attempting to show that phoneme perception is discontinuous (categorical) or continuous depending on whether the appropriate articu-

lations are discontinuous or continuous. For example, an important acoustic cue for the perception of the voiced stops /b/, /d/, and /g/, is the direction and extent of the second-formant transition. When a series of synthetic speech sounds, varied in small steps along this continuum, is presented to listeners who are asked to identify the sounds as /b/, /d/, or /g/, their responses normally tend to divide the acoustic continuum into three sharply-defined phoneme categories (Liberman et. al., 1957). The shifts from one response or phoneme label to another are very abrupt. On the other hand, when listeners are asked to identify synthetic vowels, that vary in small steps along an appropriate acoustic continuum, their responses do not divide the continuum into sharply defined phonemic categories: the shift from one category to another is gradual (Fry et al., 1962). The voiced stops in the first example are produced by very different articulations; for example, /b/ is produced by a movement of the lips and /d/ is produced by a movement of the tongue. There are no intermediate articulations between /b/ and /d/. In the production of vowel sounds, however, the articulators can move continuously from one vowel phoneme to another. Thus, the theory maintains that the perception of most consonants is discontinuous because they covertly evoke discontinuous articulations whose sensory feedback controls the overt identifications. Accordingly, the perception of vowels is continuous because the mediating articulatory movements are continuous.

The statement that perception is categorical for consonants and continuous for vowels is argumentative rather than factual. It is based on subjective evaluations of the "steepness," "sharpness," or "abruptness" of phoneme-labelling gradients relative to different

synthetic speech continua; since such comparisons are not dimensionally homogeneous the continuous-discontinuous distinction is arbitrary (Lane, 1963). With appropriate conditioning techniques, however, it is possible to transfer the stimulus control of the two kinds of phonemic contrasts to a non-speech continuum. Direct comparison of consonant and vowel identification functions would then be meaningful, since they would be relative to the same coordinate system. If the properties of speech perception are preserved in this transfer -- that is, if perception is still categorical -- then, under the hypothesis that categorical perception is effected by discretely-different mediating response, the slopes of the respective identifications functions should be steeper for consonant contrasts than for vowel contrasts.

In the present experiment, a consonant contrast (/b/, /g/) or a vowel contrast (/a/, /i/) was conditioned to a pair of visual stimuli. Measures of response frequency and latency, were then obtained at intermediate stimulus values, using procedures that were as comparable as possible to those employed by Liberman and his collaborators.

Method

Experiment I.

Subjects

Twenty paid volunteers, 11 male and 9 female undergraduates, served individually in one-hour sessions.

Apparatus

The subject (S) wore a dynamic throat microphone (Danavox) while seated, in a sound-insulated chamber, in front of a projection screen and a coin dispenser. A slide projector (Leitz VIII S), located behind S and outside of the chamber, projected the visual stimuli through

a window and onto the screen. A vocal response to the stimulus triggered a voice-operated relay, which closed an electrically-operated shutter (rise-decay time, 40 msec), advanced the projector to the next slide, and opened the shutter two seconds later. The latency of the response was measured in msec on a frequency counter (Hewlett-Packard 523D), which started when the shutter was signalled to open and stopped when it was signalled to close. The form of the response was monitored and recorded by the experimenter.

Stimuli

Nine circles, three inches in diameter, were cut out of a piece of flat black art paper. A sector, symmetrical about the radius at 360° , was cut out of each circle; the sectors that were deleted ranged from 30° to 62° in steps of 4° . The nine sectors remaining were pasted on pieces of white paper and transparencies were prepared with a Polaroid-Land copymaker (Model 208) and film (Type 146-L). The stimuli had a diameter of 10 in. when projected on the screen, 5 in. above eye level and 24 in. in front of S. Eight irrelevant stimuli, also prepared and presented in this manner, were used in identification training. These stimuli consisted of various geometrical patterns, based on variants of the cross, rhombic, polygon and solar designs.

Procedure

The 20 Ss were assigned randomly to two groups of 10 each. Those in group V (vowel contrast) were conditioned to identify the extreme stimuli in the series, numbers 1 and 9, as /b[^]b/ and /bib/, respectively; Ss in group C (consonant contrast) were conditioned to call them /b[^]b/ and /g[^]g/. Thus, the identification responses for group V were drawn from an articulatory continuum, while those for the other group (C) were "essentially discontinuous."²

Identification Training. These instructions were read to S:
(the alternative words, in parentheses, were used with group C).

"On this screen will appear various symbols. Some are /bAb/ symbols and some are /bib/ (/gAg/) symbols; still other symbols are irrelevant. Your task is to learn to identify the /bAb/ symbols and the /bib/ (/gAg/) symbols when they appear, by saying /bAb/ or /bib/ (/gAg/). Each time a /bAb/ or /bib/ (/gAg/) symbol is presented and you correctly identify it, you will get a nickel, which you may keep. If you respond incorrectly to one of the /bAb/ or /bib/ (/gAg/) symbols, or if you respond to any of the irrelevant symbols, no nickel will be dispensed."

Stimuli 1 and 9 were presented 16 times each and the eight irrelevant stimuli 5 times each, all in a randomized order. Occurrences of the /bAb/ response within three sec after the presentation of stimulus 1 were reinforced, while those of the /bib/ (or /gAg/) response were not. The relations were reversed for stimulus 9. The stimulus duration equalled the response latency or three sec, whichever occurred first. All Ss met the training criterion of ten consecutive reinforced responses within the series of 72 stimuli.

Identification test. These instructions were read to S:

"In this part of the experiment we will present only /bAb/ and /bib/ (/gAg/) symbols. Your task is to indicate whether the symbol presented each time is an example of a /bAb/ symbol or /bib/ (/gAg/) symbol. We are going to disconnect the nickel dispenser, so you will not know whether your response is correct, but I will be keeping score and

I will tell you how well you did when we are through. Be sure to respond to each presentation--guess if necessary."

The nine stimuli were presented, ten times each, in a randomized order. The stimulus duration equalled the response latency of S, while the interstimulus interval was 2 sec.

Experiment II.

The procedure of Experiment I was replicated on seven consecutive days with each of two female and two male paid Ss. The same stimuli and the same responses were used: two of the Ss were assigned the vowel contrast and two were assigned the consonant contrast.

Results

Identification and latency functions of individual Ss from Exp. I. are shown graphically in Fig. 1. The identification functions represent the number of times each stimulus was identified as a /bab/ symbol over the ten presentations of that stimulus. The /bib/ or /gag/ identification functions are complementary to those presented and are not shown. The Ss were numbered relative to the order in which they served in the experiment but their identification functions are ordered, from left to right, roughly on the basis of where each gradient and its complement intersected on the stimulus continuum. For all Ss, the response gradients are flat at one or both extremes and over an extended range of the stimulus continuum. Thus, the gradients may be described as effectively partitioning the stimulus set into two classes; the stimuli in each class are equally effective in evoking a given identification response.

Response latencies, measured as the time elapsing from the onset of the stimulus to its termination by either response, are represented in Fig. 1 (dashed lines) by their median value at each stimulus.

Although each determination is based on only ten responses, the resulting functions are orderly for most Ss and show a close relation to the corresponding identification function: (1) maximum latencies occur at or near the point where the pair of identification responses are equiprobable; (2) latencies are approximately equal over the stimulus ranges corresponding to plateaus in the identification function.

Results based on repeated measurements over several consecutive days of testing for the four Ss in Experiment II were in accord with the results reported above. Response gradients for two of these Ss, one for each experimental condition, are shown in Figs. 2 and 3. The identification functions for these Ss are based on the last six days of testing. Each data point represents, for a total of 60 stimulus presentations, the proportionate number of times that stimulus was identified as a /bAb/ symbol. Response latencies are represented by modified bar histograms which show how the latencies were distributed with respect to ten equal class intervals on a logarithmic scale. The approximate midpoint of each class interval is indicated on the latency scales in Figs. 2 and 3. The median latency for each distribution is indicated by a filled triangle. The identification functions for these two Ss are nearly identical and their latencies reveal the same general functional relation to response probability. Response latencies for S 22 are longer overall than those for S 23 and show greater variability, but the response latencies of both Ss are distributed over their minimum values when the probability of response is either zero or 1.00 and they are distributed over their maximum values when the probability of response is nearest .50.

In Fig. 4, the results obtained in Experiment I are compared with those obtained with synthetic speech stimuli (/do/ - /to/) in a study by Liberman et al., (1961a). Since different stimulus continua were involved in these two studies a direct comparison of the respective identification functions would be inappropriate. Therefore, comparison is based on the following procedure. For each S, each stimulus was compared to every other in the percent of /bAb/ responses it evoked, in the case of the present study, or in the percent of /do/ responses evoked, in the case of the synthetic speech study. The difference in percent was computed for each comparable pair. A frequency distribution of these percentage differences was prepared, with class intervals of 10 percent, converted to relative frequencies and plotted in Fig. 4. If the individual gradients were step-functions, the limiting case of "categorical perception," then all the percent differences would be either zero or 100, and the frequency distribution would pile up in the extreme classes leaving intermediate classes empty. On the other hand, if perception was continuous, the case where response probability is a strictly monotonic function of stimulus change - then the above frequency distribution would be more nearly rectangular in shape. Figure 4 shows that the distributions for groups V and C do not differ systematically either from each other or from that based on synthetic speech stimuli.

Discussion

The above findings demonstrate that stimulus control of phonemic contrasts is readily transferred to a non-speech continuum and that the properties of stimulus identification formerly attributed exclusively to speech perception are preserved in this transfer.

The identification functions obtained in this study show that response probability remains constant at zero or unity over wide ranges of stimulus variation. As shown in Fig. 4, the distribution of differences in percent identification of visual stimuli does not differ appreciably from that of speech stimuli, although the set of speech identification gradients selected for comparison is the most "categorical" of those reported in the literature.

The response-latency functions obtained in this study, with both vowel-contrast and consonant-contrast labelling responses, reach a maximum at the "boundary" of their corresponding identification gradients and a minimum within the plateaus of those gradients. These findings for response latency in identification of non-linguistic visual forms also accord with those reported for identification of synthetic speech stimuli: "For both stops and vowels, voice and button-press reaction times were lower near phoneme centers than in the vicinity of phoneme boundaries," (Studdert-Kennedy, Liberman, and Stevens, 1963).

Advocates of the motor theory of speech perception interpret the properties of the identification and latency gradients described above as evidence that perception is categorical and is mediated by covert articulation (or its neural correlates). If we conclude, therefore, that 20 minutes of discrimination training sufficed to transfer the mediation processes of speech perception to a visual continuum, then we must ask why the consonant contrasts did not yield more categorical gradients of identification and latency than the vowel contrasts. If categorical perception is effected by discretely different mediating responses, then the consonant-contrasts should have been much sharper than the vowel contrasts. If it is maintained, on the other hand,

that the mediation hypothesis is applicable only to the perception of speech continua and is predicated on extensive verbal conditioning (as the motor theorists would have it), then we must ask why the identification and latency functions for visual symbols so closely resemble those for speech signals. Since non-linguistic visual forms and synthetic speech stimuli both effect "categorical perception", and since mediating articulation is a gratuitous assumption in the former case, it seems unwarranted in the latter.

References

- Fry, D. B., Abramson, A. S., Eimas, P. D. & Liberman, A. M. The identification and discrimination of synthetic vowels. Lang. Sp., 1962, 5, 171-189.
- Lane, H. L. The motor theory of speech perception: a critical review. In: Experimental analysis of the control of speech production and perception: IV. Behavior Analysis Laboratory, University of Michigan, 1963.
- Liberman, A. M., Cooper, F. S., Harris, Katherine S. & MacNeilage, P. F. A motor theory of speech perception. Proced. Speech Commun. Seminar, Royal Institute of Technology, Stockholm, 1963.
- Liberman, A. M., Harris, Katherine S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol., 1957, 54, 358-368.
- Liberman, A. M., Harris, Katherine S., Kinney, JoAnn, & Lane, H. L. The discrimination of relative onset time of the components of certain speech and non-speech patterns. J. exp. Psychol., 1961, 61, 379-388. (a)
- Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. Paper read at the 66th meeting of The Acoustical Soc. Amer., November, 1963.

Footnotes

1. This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education.
2. Vide Liberman et al., 1957.

Figure Captions

Fig. 1. Gradients of response frequency and latency for individual Ss in Experiment I. The points connected by continuous lines represent the number of /bAb/ responses evoked by the 10 presentations of each stimulus. Complementary functions were obtained for the /bib/ and /gAg/ responses and are not shown. The points connected by broken lines represent the median latencies of all responses emitted to the corresponding stimulus. Latencies were not measured for S 12.

Fig. 2. Identification and response latency for one S from experiment II. The relative frequency of the /bAb/ response, in 60 presentations of the corresponding stimulus on each of six consecutive days, is shown by the dashed lines. Frequency distributions of the latencies of the /bAb/ and /bib/ responses combined are shown at each stimulus value; the median latency is represented by the filled triangle.

Fig. 3. Identification and response latency for one S from experiment II. The relative frequency of the /bAb/ response, in 60 presentations of the corresponding stimulus on each of six consecutive days, is shown by the dashed lines. Frequency distributions of the latencies of the /bAb/ and /gAg/ responses combined are shown at each stimulus value; the median latency is represented by the filled triangle.

Fig. 4. A comparison of the degree of categorical perception obtained with identification of synthetic speech stimuli and identification of non-linguistic visual forms. The triangles show the transformed labelling gradients for the phonemic contrast /do/ - /to/, reported in Liberman et al., (1961a). The open and filled squares represent the transformed labelling gradients shown on the top and bottom of Fig. 1, respectively, for identification of non-linguistic visual forms.

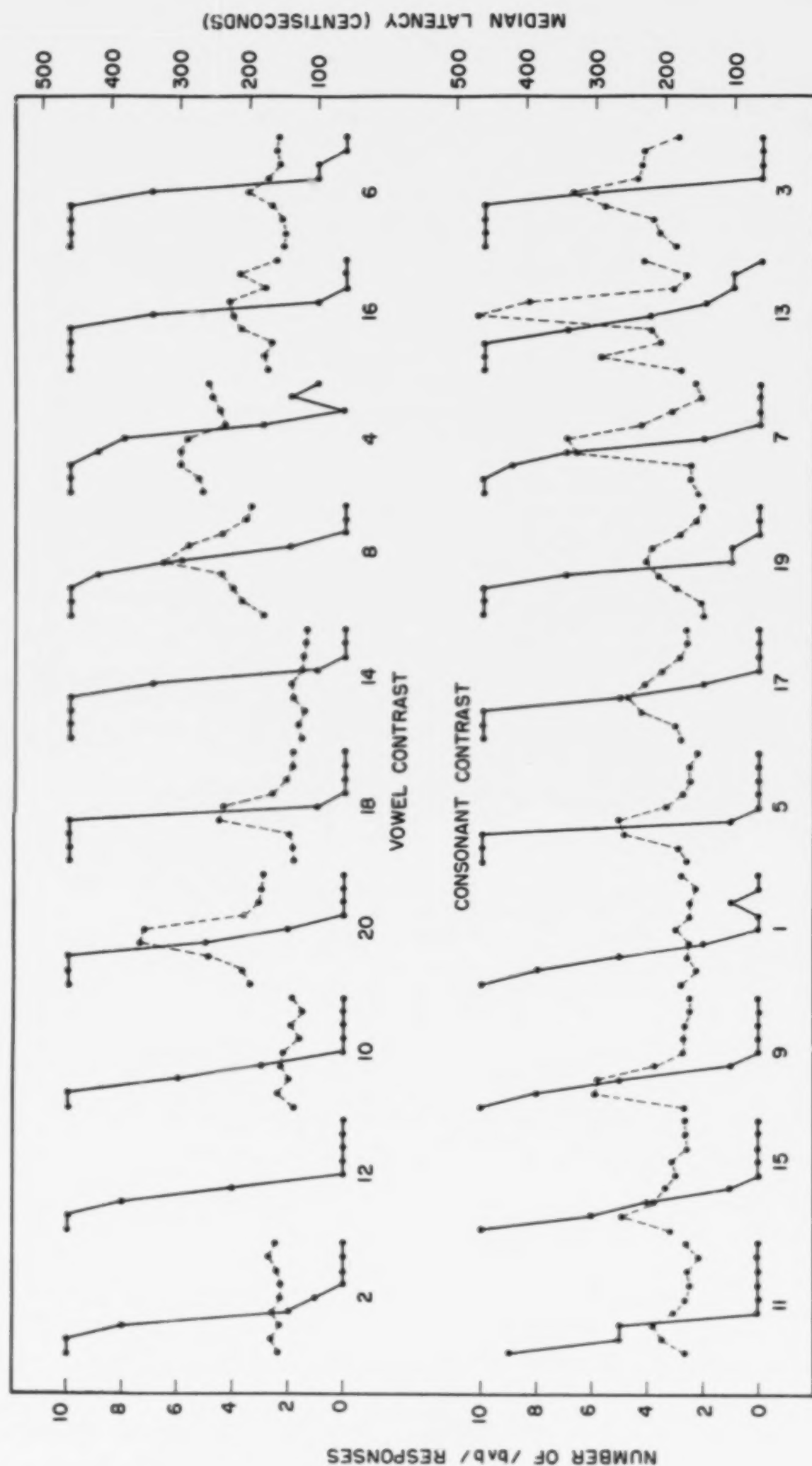


Fig. 1

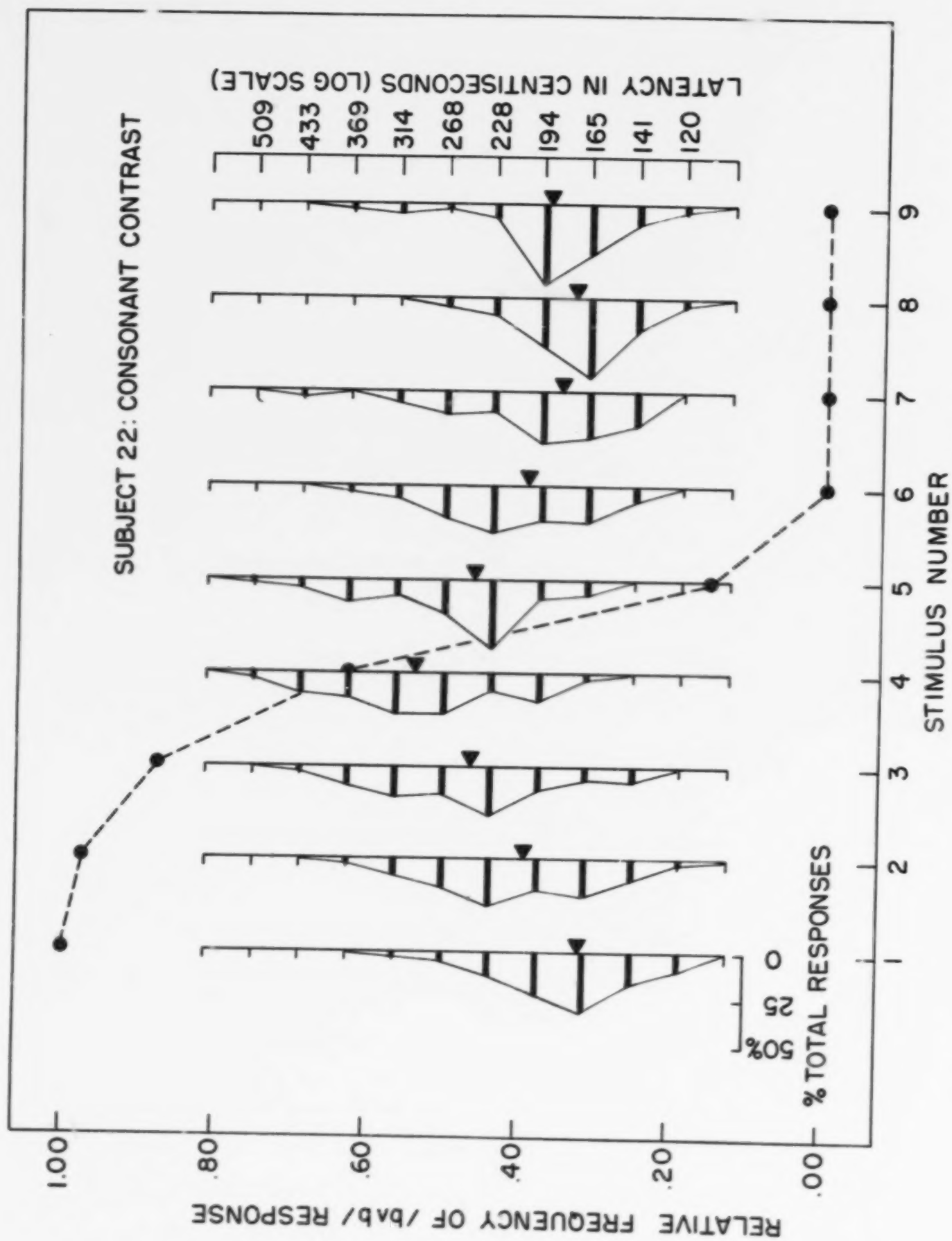


Fig. 2

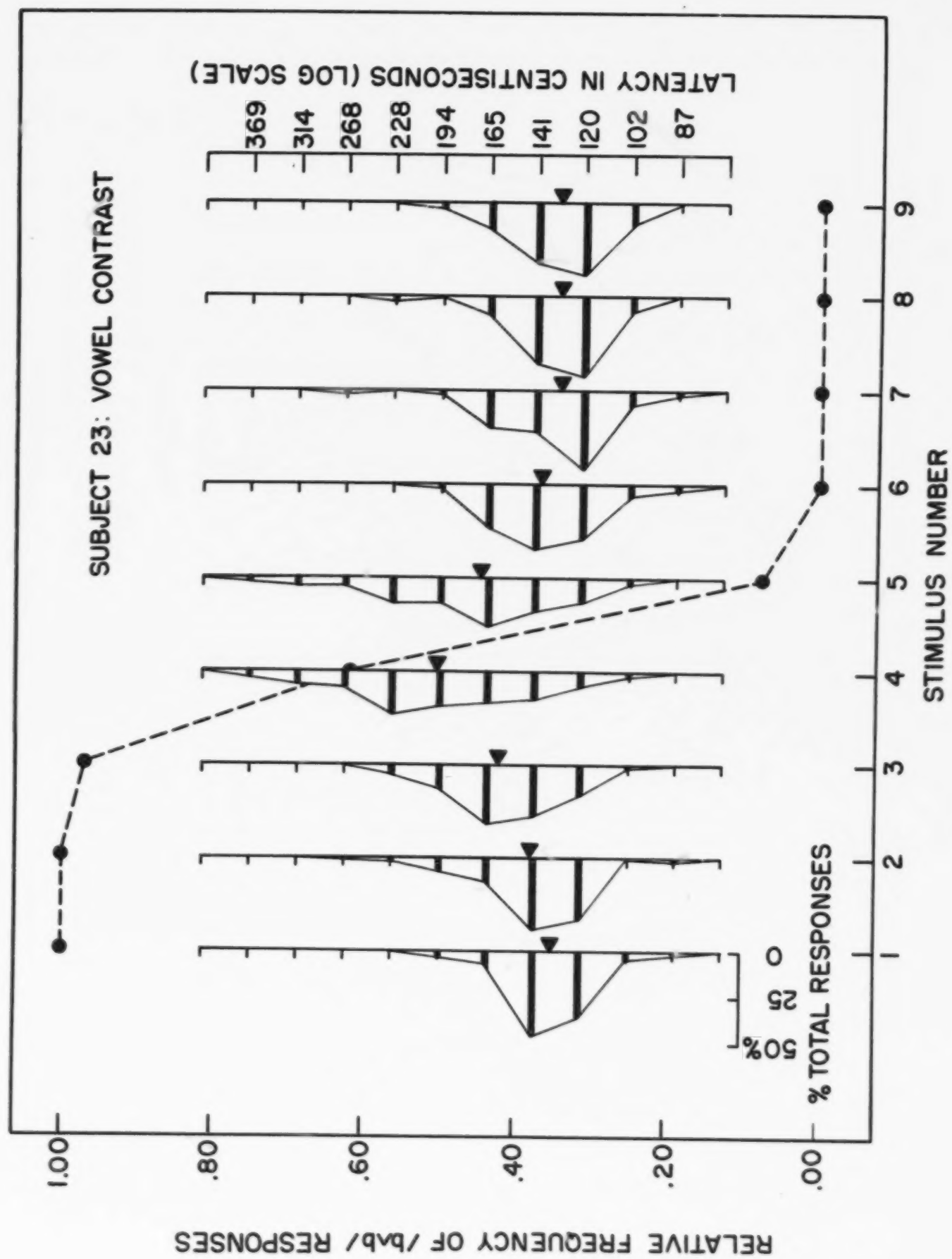


Fig. 3

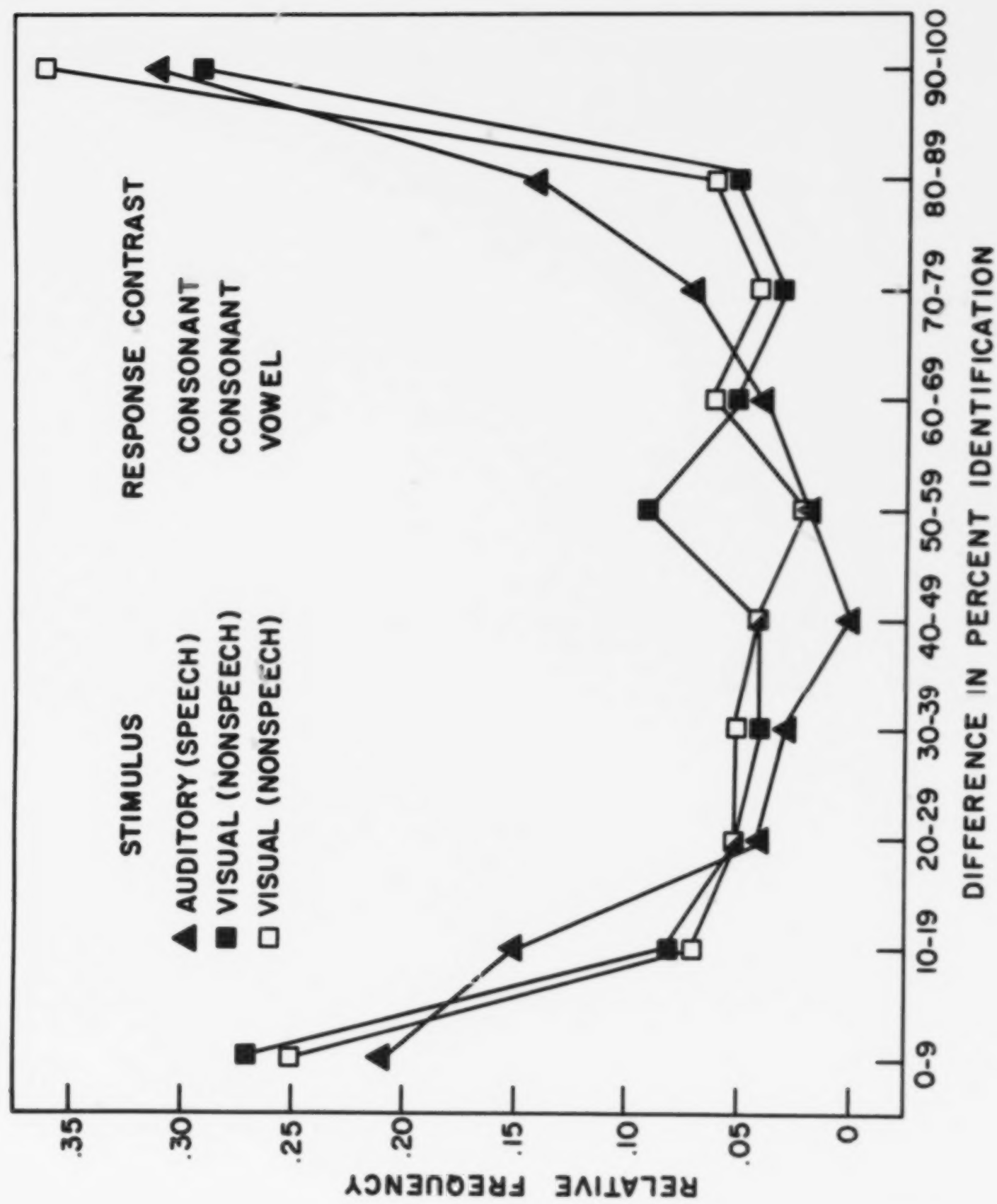


Fig. 4

ATTENTION TO SINGLE STIMULUS PROPERTIES IN THE
IDENTIFICATION OF COMPLEX TONES

David V. Cross and Harlan Lane
The University of Michigan

Attention to Single Stimulus Properties in the
Identification of Complex Tones.

David Cross and Harlan Lane
Behavior Analysis Laboratory
University of Michigan

In the classical paired-associate paradigm, and in identification-learning tasks in general, the experimenter prescribes an identification function, i.e., a rule that assigns a particular response from a set of possible responses to a single stimulus from an experimentally-defined stimulus set. Typically, an equal number of stimuli and responses are placed in a one-to-one correspondence and reinforcement is contingent on emitting the prescribed response in the presence of a given stimulus. When the stimuli differ along a single physical continuum, it is normally possible to present additional stimuli having values on the continuum intermediate to the initial training stimuli, in order to assess how the various responses generalize to other stimuli. Previous investigations, utilizing unidimensional stimulus sets, have revealed that subjects' responses partition the continuum into response classes (Cross and Lane, 1962; LaBerge, 1961; Liberman et al., 1957). Stimuli within a class are equally effective in evoking a given response, and a class boundary can be defined between adjacent classes at the point on the stimulus continuum where response probabilities are equal. For convenience we will refer to this effect of identification training by saying that the prescribed identification functions induce an identification structure on the stimulus set which is revealed in standard generalization-testing procedures. Properties of identification structures induced on unidimensional stimulus sets have been described in detail by Lane (1963).

If the stimuli in an identification-learning task differ on more than one physical dimension, what identification structure will be induced on the multidimensional stimulus set? In this case, how do differences along component dimensions combine to determine the resulting identification structure? When the identification function for a multidimensional stimulus set is prescribed for certain combinations of stimulus properties, but correct identifications can be made by selectively attending single component properties, which properties will be attended, and what identification structure will be induced on the set? The present study was designed to assess the properties of multidimensional stimulus control in identification experiments as a function of the way stimulus properties are combined in the prescribed identification function.

Method

Stimuli

The stimuli were complex tones varying in both fundamental frequency and harmonic structure. They were synthesized in the following manner. The primary signal source was a pulse generator (General Radio 1217-B), externally triggered by an audio oscillator (Hewlett-Packard 241A). The spectral distribution of the pulse trains comprised a fundamental frequency (f_0), determined by the oscillator setting, and all of the harmonics of f_0 . These pulses, one msec. in duration, were passed through a variable bandpass filter (Khron-hite 310 AB) that was set to attenuate all harmonics (at 24 db/octave) except those occurring within a designated range, the filter bandpass, which was varied. Each stimulus was characterized, therefore, by two variable parameters: its fundamental frequency (f_0) and the center frequency (F_1) of the filter bandpass.

A total of 25 different stimuli were obtained by combining five levels of f_0 (110, 130, 150, 170, and 190 cps) and five levels of F_1 (514, 819, 1192, 1645 and 2160 cps). The values for F_1 corresponded to the following filter settings: 394-670, 670-1000, 1000-1420, 1420-1900, and 1900-2450 cps. These values marked off equal intervals on the mel scale. The stimuli, 3 sec in duration, were pre-recorded at 8 sec intervals and at constant VU level on magnetic tape (Ampex 351-2U). Throughout the rest of this report the following notation will be used in reference to the above stimuli: $S_{j,k}$ will denote the stimulus resulting from combining the j^{th} level on f_0 with the k^{th} level on F_1 ; $j = 1, 2, 3, 4, 5$ and $k = 1, 2, 3, 4, 5$. Thus $S_{3,2}$ represents a tone having $f_0 = 150$ cps and $F_1 = 819$ cps.

Apparatus

The subject (S) was seated in a sound-attenuating room, in front of a rack-mounted panel and shelf. A small tray, on the front of the panel, received nickels from an electrically-operated coin dispenser, mounted behind the panel. The operanda were five push-button switches mounted in a 14" x 8" x 8" minibox, which was attached to the shelf beneath the coin tray. The actuating mechanisms for the switches were 1" x 1 1/4" colored, translucent-plastic buttons, that could be trans-illuminated by miniature lamps. The buttons were colored differently and arranged in a semicircular array in order to minimize response induction.

The stimuli were reproduced on a tape recorder and presented to S binaurally through calibrated earphones (Grason-Stadler TDH-39). An electronic switch (Grason-Stadler 829D), interposed between the tape recorder and the earphones, closed upon presentation of the stimulus

and opened when a response button was pressed, thus terminating the stimulus. Stimulus presentation, reinforcement and other experimental events were all controlled by sensitive relays, stepping relays, and predetermining counters, which were operated, in turn, by coding tones, synchronized with the stimuli and pre-recorded on a second track of the magnetic tape.

Procedure

Twelve college students served individually in single sessions lasting approximately one hour. An experimental session involved two phases: a training phase, during which S was given an identification task requiring the press of a different button in response to each of five different stimuli, and a test phase, during which generalization along the component stimulus dimensions could be assessed as a function of the training condition. Two training conditions prevailed, differing only in the stimuli that were presented. The six Ss in Group I were required to emit responses R_1 through R_5 , corresponding to the lateral position of the response button, according to the following identification functions, i.e., stimulus-response assignments: $S_{1,1} \rightarrow R_1$, $S_{2,2} \rightarrow R_2$, $S_{3,3} \rightarrow R_3$, $S_{4,4} \rightarrow R_4$ and $S_{5,5} \rightarrow R_5$. The six Ss in Group II were presented a different set of stimuli and were required to learn the following identification functions: $S_{1,5} \rightarrow R_1$, $S_{2,4} \rightarrow R_2$, $S_{3,3} \rightarrow R_3$, $S_{4,2} \rightarrow R_4$ and $S_{5,1} \rightarrow R_5$. Thus, for Group I the training stimuli increased in F_1 directly as they increased in f_0 . For Group II, the training stimuli decreased in F_1 as they increased in f_0 . Although Group I Ss were presented stimuli along the positive diagonal of the stimulus matrix and Group II Ss were presented stimuli along the negative diagonal, Ss from the two groups received identical training.

Training. The following instructions were read to S:

"You can earn money in this experiment by learning to correctly identify each of five sounds that will be presented to you through these earphones. To each of the five sounds there corresponds a different button on this box. Each time a sound is presented you are to press one of these buttons. If you press the right one, it will light up. If you press the wrong button, the one you should have pressed will light up. You will earn a nickel, which will be immediately dispensed, only when you make five correct identifications in a row. An error cancels the correct responses that preceded it, and only runs of five correct responses will be paid off. Your response, right or wrong, will terminate the sound, so you can respond only once to each presentation. Each presentation will last 3 secs. Failure to respond during this period will be counted as an error so you must respond every time. You will not earn much at first because you will have to guess what button goes with what sound. The faster you learn, however, and the fewer errors you make, the more money you will earn."

The five training stimuli were presented to S, 40 times each, in a pre-recorded, randomized sequence. If S was still making errors in identification at the end of this sequence, it was started again and continued until 25 consecutive correct identifications were obtained.

Testing. The experimenter reentered the chamber and told S that in this phase the same five sounds would be presented and he was to continue responding in the same way, but all feedback regarding the

correctness of his responses would be discontinued: the buttons would no longer light up and the coin dispenser would be inoperative. Correct identifications would still earn money, now at the rate of one nickel each. The subject was also told that, in addition to the five training stimuli, a few additional stimuli would be randomly interspersed in the series of presentations; he was told to respond to these novel stimuli as he would have responded to them had they occurred during the previous phase. The subject was instructed to respond to every presentation; each failure to respond would subtract one nickel from his earnings. The test stimuli were presented to S in a pre-recorded sequence, composed of 10 random permutations of the original 25 stimuli, or a total of 250 stimulus presentations.

Results

Figures 1 and 2 show, for Groups I and II respectively, how the identifications established during training generalized to other pairwise combinations of stimulus levels. The discrete distributions in these figures represent, for each stimulus, the relative number of times each of the five responses was emitted in the presence of the designated stimulus. Each determination was based upon 60 presentations of $S_{j,k}$ (10 to each S). Bar heights denote the proportion of $S_{j,k}$ presentations resulting in R_1 , which may be denoted $P(R_1|S_{j,k})$ and interpreted as empirical estimates of the conditional response probabilities. Attention is specifically directed to the orderly manner in which these response distributions either change or remain constant in form from one stimulus level to the next. For Group I (Fig. 1), the modal response for each distribution is constant over levels of F_1 but changes systematically over levels of f_0 . For Group II (Fig. 2) the opposite result was

obtained: the modal response is constant over levels of f_0 (with only one exception at $S_{1,2}$) and changes over levels of F_1 . It should be noted that, although this result is reported in terms of pooled data, it holds for individual $\underline{S}$ s: as $S_{j,k}$ was changed over different levels of both f_0 and F_1 , each $\underline{S}$ in Group I responded according to changes in f_0 only, ignoring F_1 , and each $\underline{S}$ in Group II attended F_1 , in disregard of changes in f_0 .

An alternate and, in some respects, more convenient representation of these results is possible since the responses were physically ordered from left to right. The $\underline{S}$'s task in this experiment was formally analogous to a category-scaling task, in that five categories were used for rating stimuli. In the same way that a mean category assignment for each stimulus may be calculated, we may define a response measure for each $S_{j,k}$, denoted $M(S_{j,k})$, which is based on the conditional probability of R_i times a numerical assignment to each R , $N(R_i)$, that preserves the natural response order. Thus, $M(S_{j,k}) = \sum_i N(R_i) \cdot P(R_i | S_{j,k})$. The $N(R_i)$ in this case are simply numbers used to label the responses. The measure $M(S_{j,k})$ induces a partial order on the stimulus set which is preserved under any monotone increasing transformation of the $N(R_i)$. The graphs in Fig. 3 result from assigning the numbers 1 through 5 to R_1 to R_5 respectively and calculating $M(S_{j,k})$ for each stimulus, as described above. The left-hand graphs in Fig. 3 depict $M(S_{j,k})$ as a function of f_0 with levels of F_1 as a parameter, and the right-hand graphs depict the same $M(S_{j,k})$ as a function of F_1 with levels of f_0 as a parameter. Comparison of Groups I and II (top vs. bottom graphs) is more convenient in this representation than it is in Figs. 1 and 2. The significant feature of this comparison is that, for Group I, the levels

of f_0 were simply ordered by the function $M(S_{j,k})$ and the levels of F_1 were not; for Group II, the opposite effect obtained. Thus, the only relevant stimulus property exerting behavioral control was the fundamental frequency, f_0 , for Group I, and the resonant frequency, F_1 , for Group II.

Discussion

The fact that Ss may selectively attend one stimulus property and not others in discrimination situations involving multidimensional stimuli is not a novel observation (vide Cross, 1963; Sheppard, 1962; Reynolds, 1961; Warren, 1953, 1954; Harlow, 1945). The question which we now wish to consider is: why does the training of differential responding to stimuli differing systematically according to one combination of stimulus properties result in attention directed to one dimension, whereas training involving a different combination of the same stimulus properties results in attention to a different dimension? It seems reasonable to conjecture that when appropriate discriminative behavior is possible by attending only stimulus variations along a single dimension, regardless of what other relevant differences exist, a subject will attend the dimension along which the differences are most noticeable, i.e., most readily observed. There is evidence to support this notion (Warren, 1953, 1954; Harlow, 1945). Considering the outcome of the present study, we might infer that the training stimuli for Group I differed more in apparent pitch, f_0 , than in timbre, the probable psychological correlate of F_1 . Thus, differences in f_0 were attended and differences in F_1 ignored during training; f_0 then persisted as the controlling stimulus dimension during generalization testing. For the training stimuli presented to Group II, we might infer that timbre

differences dominated pitch differences, providing the converse result during testing. These inferences can be tested empirically by repeating this experiment with different sets of training stimuli, obtained by rotating the physical coordinate system, until the present results fail to obtain. An alternative analysis suggests that the physical property of the stimulus actually exercising behavioral control was not F_1 , but was instead the number of harmonics of f_0 actually present in the acoustic signal. Since the width of the bandpass was constant for a given level of F_1 , the number of audible harmonics depended upon f_0 . The difference between the number of harmonics present in $S_{1,5}$ and that in $S_{5,1}$ was greater than the difference between that present in $S_{1,1}$ and in $S_{5,5}$. If the number of harmonics present was, in fact, the controlling property for Group II, then a stimulus set constructed so that the number of audible harmonics remained constant as f_0 was varied across each level of F_1 (obtained by varying the bandwidth around F_1) would result in an outcome different from that found in this study.

The converging operations suggested above are intended to answer the question of what are the actual dimensions of behavioral control for acoustic stimuli characterized by the parameters of fundamental frequency and harmonic structure. Interest in these properties of auditory stimuli extends beyond the limits of the present study, since the perception of speech sounds, in particular the vowels, involves the same stimulus properties. Thus, specification of the parameters of behavioral control for complex acoustic stimuli would enhance

our understanding of the variables involved in speech recognition and our ability to design effective programs for language learning.

Figure Captions

Fig. 1. Group I. Each histogram depicts the proportionate number of R_1 through R_5 responses emitted to each stimulus, in ten presentations to each of 6 S_s .

Fig. 2. Group II. Each histogram depicts the proportionate number of R_1 through R_5 responses emitted to each stimulus, in ten presentations to each of 6 S_s .

Fig. 3. Ordinal measures $M(S_{j,k})$ computed for each stimulus as a function of f_0 (left side) with levels of F_1 as a parameter, and as a function of F_1 (right side) with levels of f_0 as a parameter. Results are presented for both Group I (top figures) and Group II (bottom figures). Only the relative ordering of the stimuli with respect to the vertical axis, is relevant. Filled circles denote response measures for stimulus values employed during training.

References

- Cross, D. V. Metric properties of multidimensional stimulus generalization.
In: Experimental analysis of speech production and perception: III.
Unpublished report, Behavior Analysis Laboratory, University of
Michigan, 1963.
- Cross, D. V. & Lane, H. L. On the discriminative control of concurrent
responses: the relations among response frequency, latency, and
topography in auditory generalization. J. exp. Anal. Behav.,
1962, 5, 487-496.
- Harlow, H. F. Studies in discrimination learning by monkeys: VI.
Discrimination between stimuli differing in both color and form,
only in color, and only in form. J. gen. Psychol., 1945, 33,
225-235.
- LaBerge, D. Generalization gradients in a discrimination situation.
J. exp. Psychol., 1961, 62, 88-94.
- Lane, H. L. The motor theory of speech perception: a critical review.
In: Experimental analysis of speech production and perception: IV.
Unpublished report, Behavior Analysis Laboratory, University of
Michigan, 1963.
- Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. The
discrimination of speech sounds within and across phoneme boundaries.
J. exp. Psychol., 1957, 54, 358-368.
- Reynolds, G. S. Attention in the pigeon. J. exp. Anal. Behav., 1961,
4, 203-208.
- Sheppard, R. N. Attention and the metric structure of the stimulus
space, 1962. Unpublished report, Bell Telephone Laboratories,
Murray Hill, New Jersey.

Warren, J. M. Additivity of cues in a visual pattern discrimination

by monkeys. J. comp. physiol. Psychol., 1953, 46, 484-486.

Warren, J. M. Perceptual dominance in discrimination learning by

monkeys. J. comp. physiol. Psychol., 1954, 47, 290-293.

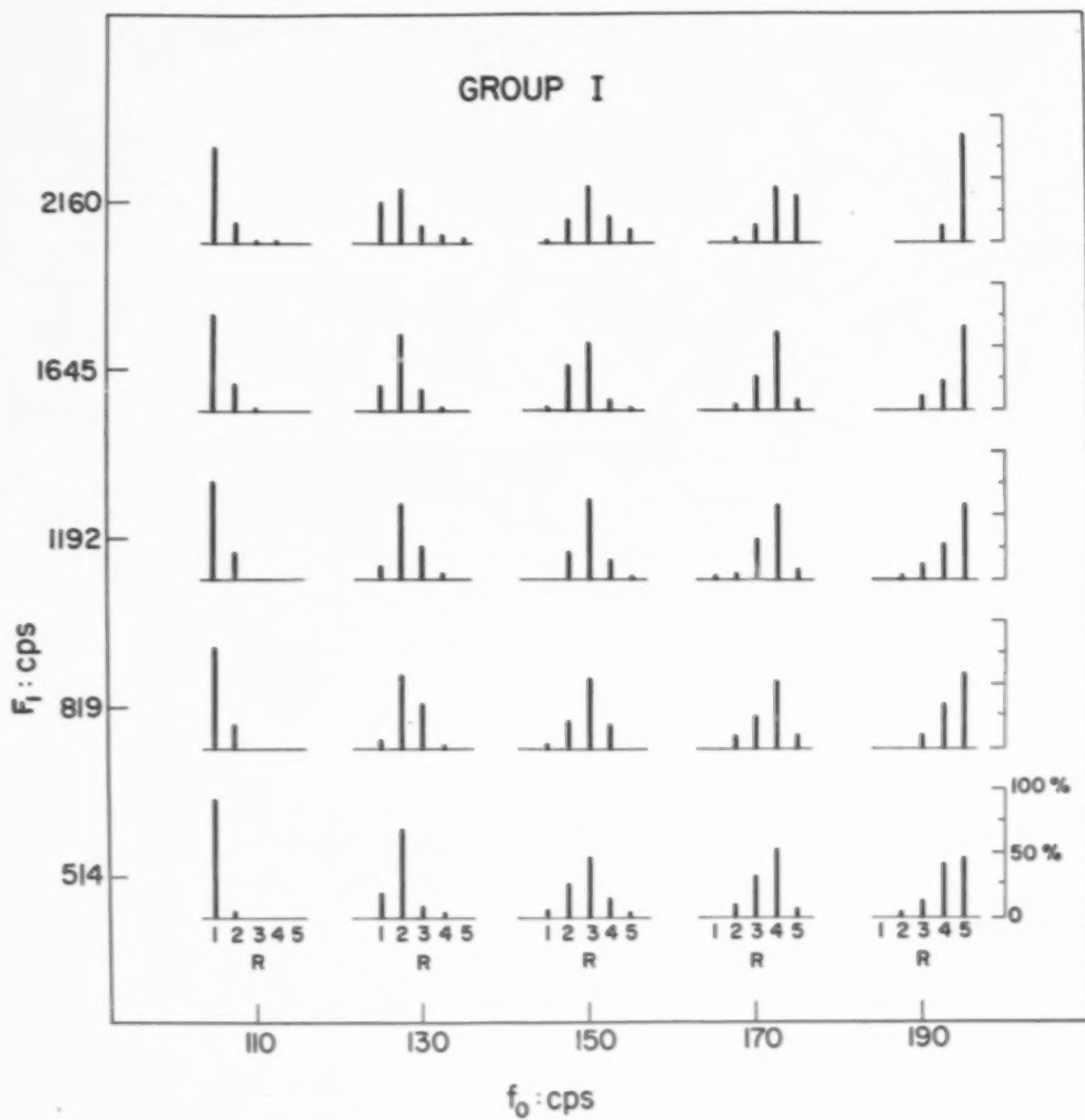


Fig. 1

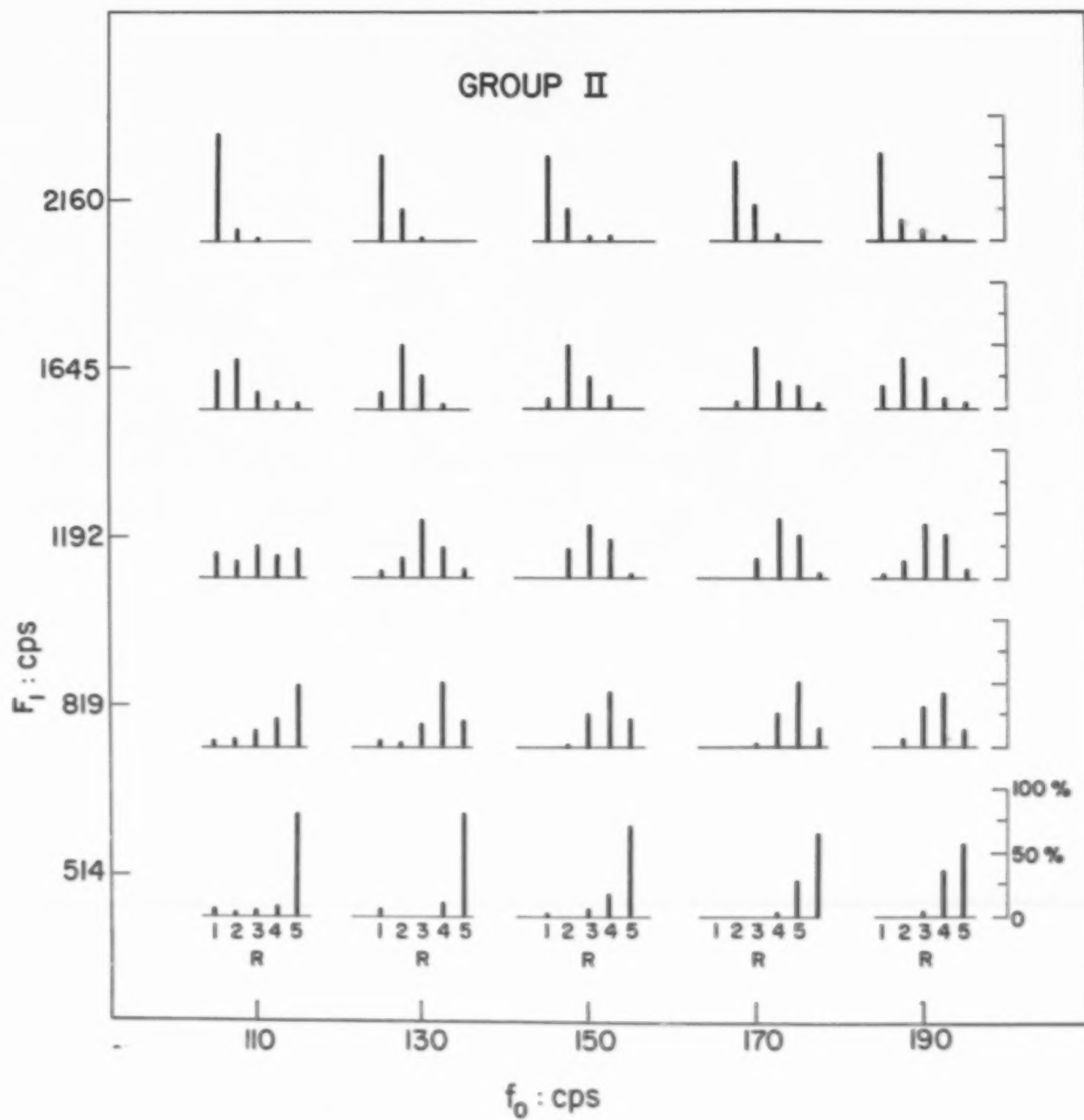


Fig. 2

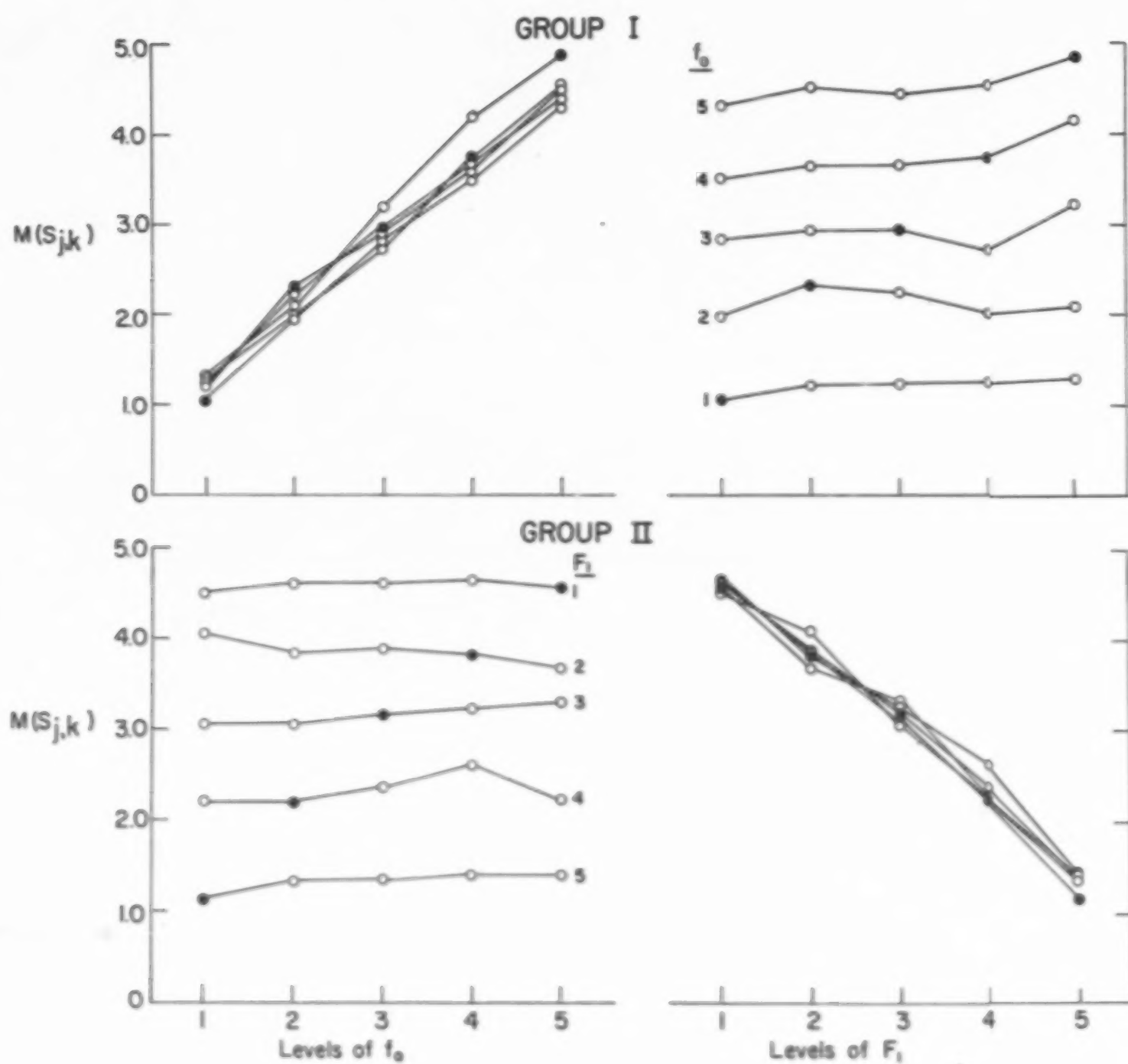


Fig. 3

A TECHNIQUE FOR PRODUCING KNOWN SOUND PRESSURES AT THE EARDRUM

Strange Ross
The University of Michigan

A Technique for Producing Known Sound

Pressures at the Eardrum¹

Strange Ross

Behavior Analysis Laboratory

University of Michigan

When attempting to specify the intensity parameter of pure-tone stimuli, various difficulties are encountered, depending upon how the stimulus is presented to a listener. In a free-field situation, the physical intensity prior to stimulation is well-defined; when the subject is placed in the field, however, no well-defined parameter descriptive of the acoustical input to the hearing mechanism is readily available. In the closed-ear situation, when ordinary earphones are used, the response of the earphones may be determined by means of an artificial ear of standardized construction. However, in this latter case as well, a parameter describing the actual stimulation of a particular ear is not readily available. The reason for this is that the response of an earphone depends upon its acoustic termination, and the acoustic terminations provided by an artificial ear and a real ear differ in ways which are not readily specified.

One way in which to overcome these difficulties is to measure the sound pressure produced by a stimulating transducer when it is as close to the eardrum as possible. This can be accomplished by means of a probe microphone, consisting of a calibrated condenser microphone to which is attached a tube of suitable length and diameter. To facilitate the insertion of the probe microphone into the ear canal, to maintain its position (with its end close to the eardrum), and

also to facilitate a reproducible application of the acoustic stimulation, a moulded ear insert with provisions for these functions may be employed. The following is a detailed description of this technique.

For each individual ear, an ear insert is made from a mould, protruding as far into the ear canal as possible; (a separation of 5 - 10 mm between end of ear insert and eardrum seems to be the practical minimum). This ear insert contains a sound conducting canal, and two pieces of stainless steel tubing (supplied with the Bruel & Kjaer Probe Microphone Kit). One of the steel tubes serves to equalize pressure differences between the atmosphere and the air volume enclosed between the ear insert and the eardrum, and the other tube serves as a guide for the insertion of the probe microphone, so that the probe tip just protrudes beyond the end surface of the ear insert.

The earphone used for driving the ear insert is a small dynamic earphone (Beyer DT-508) which terminates in a cylindrical tube with a diameter of 1/4"; the earphone is connected to the sound conducting canal of the ear insert by means of a short plastic adapter. The total distance from earphone diaphragm to the tip of the ear insert is approximately 1 1/4".

The probe microphone employed consists of a Bruel & Kjaer Condenser Microphone (4134) and a probe attachment constructed from the Bruel & Kjaer Probe Microphone Kit (UA0040). The probe attachment consists of a cone which screws onto the microphone, and which supports the probe tube. A tube with diameter of 1 mm, length 47 mm, was selected. (The larger the diameter, the less attenuation is introduced by the probe tube; the smaller the diameter, the smaller the resonances introduced by the probe tube.) A small piece of 0.5 mm tubing inserted into

the free end of the probe tube reduced the resonances of this tube to a considerable degree.

The diameter of the air pressure-equalizing tube moulded into the ear insert is 0.5 mm; thus the acoustical leakage afforded by this tube is kept at a minimum. The diameter of the steel tube which guides the probe microphone is 2 mm. The length of this guide tube is 1 mm shorter than the length of that part of the probe tube which protrudes beyond the supporting cone. In this way, the probe tube protrudes 1 mm beyond the end of the guide tube, when the probe tube is inserted into the ear insert. By means of a rubber gasket and some grease a seal is produced between the probe tube and the guide tube.

The calibration of a probe microphone is usually done by means of a special coupler supplied with the Probe Microphone Kit. In this particular case, however, a different procedure must be employed. The reason for this is that when the probe microphone is mounted in the ear insert, it will pick up an airborne acoustic input from its tip as well as an acoustic input transmitted mechanically through the ear insert from the sound conducting tube to the walls of the probe tube; this latter component would not be taken into account if the regular calibration procedure were employed.

In order to take this acoustical "leakage" into account, the ear insert is mounted in a coupler containing a calibrated condenser microphone. The insert is mounted so that the tip of the probe microphone, protruding from the ear insert, is as close as possible to the diaphragm of the calibrating microphone in the coupler. The volume enclosed by the coupler is of no particular importance in this case. The response curve for the probe microphone in this particular operating

condition may be plotted by driving the coupler by the earphone which is attached to the ear insert, and then reading the outputs of the probe microphone and the calibrating microphone. (It is possible to use just one microphone for this calibration procedure. When a reading is taken from the microphone in one position, a dummy microphone supplied with the Probe Microphone Kit is substituted in the microphone's other position; and vice-versa.)

By means of the calibrated probe microphone, it is now possible to measure the sound pressure produced by the earphone and delivered through the ear insert close to the eardrum. A series of such measurements for given inputs to the earphone constitutes a real-ear calibration of the earphone-ear insert combination for the given ear.

Figure 1 shows an example of a probe microphone calibration for a particular ear insert, and, for comparison, the regular calibration of the same probe microphone by means of the coupler supplied with the Probe Microphone Kit. It will be seen that the transmission effect through the ear insert sets in at about 1000 cps, and increases in size with increasing frequency. Due to resonances in the probe tube, variability of the probe-ear insert measurements above 5000 cps occurs, which sets an upper bound to the useful frequency range.

The top curve of Fig. 2 shows the real-ear calibration with a given combination of earphone and ear insert. The lower curve shows the calibration of the same earphone-ear insert combination as obtained on a standard 2 cc coupler. By decreasing the volume enclosed by this standard coupler (by filling up the cavity with Plastalene) curves intermediate between the two curves of Fig. 2 are obtained. It is readily seen that the 2 cc coupler calibration affords a poor approx-

imation of the real ear calibration; at 4000 cps the discrepancy amounts to about 30 db. Even when the volume enclosed by the artificial ear is made as small as possible, discrepancies up to -20 db and +4 db are observed.

Below 5000 cps, the variation of the real ear calibration in repeated measurements (on different days) was found to average below ± 0.5 db, with a maximum deviation of ± 1 db. This implies that it is possible for most practical purposes to rely on the calibration curve, rather than making new measurements with the probe microphone before each stimulation. It is impractical, however, to leave the probe microphone permanently attached to the ear insert. Therefore, a dummy probe is made up which affords the same acoustical termination of the ear insert as that provided by the probe microphone, and this dummy is attached to the ear insert in place of the probe microphone. The dummy probe was constructed from a tube of the same diameter and the same length as the one used in the probe microphone, and dampened in the same way. This tube was then terminated by a tube of larger diameter, closed at the free end, so that a volume was enclosed equal to the equivalent volume of the condenser microphone attached to the real probe; (this equivalent volume is stated in the Calibration Chart for each Bruel & Kjaer Condenser Microphone). The dummy probe is held in place after insertion into the guide tube by means of a rubber stopper, which provides an airtight seal.

Footnote

1. This measurement technique was developed as the first step in obtaining accurate equal-loudness measurements under closed-ear conditions.

Figure Captions

Fig. 1. Coupler (a-a) and ear insert (b-b) calibration of probe microphone.

Fig. 2. 2 cc coupler (a-a) and real ear (b-b) calibration of a given combination of earphone and ear insert. (Curve a-a transcribed from Automatic Frequency Response Recording.)

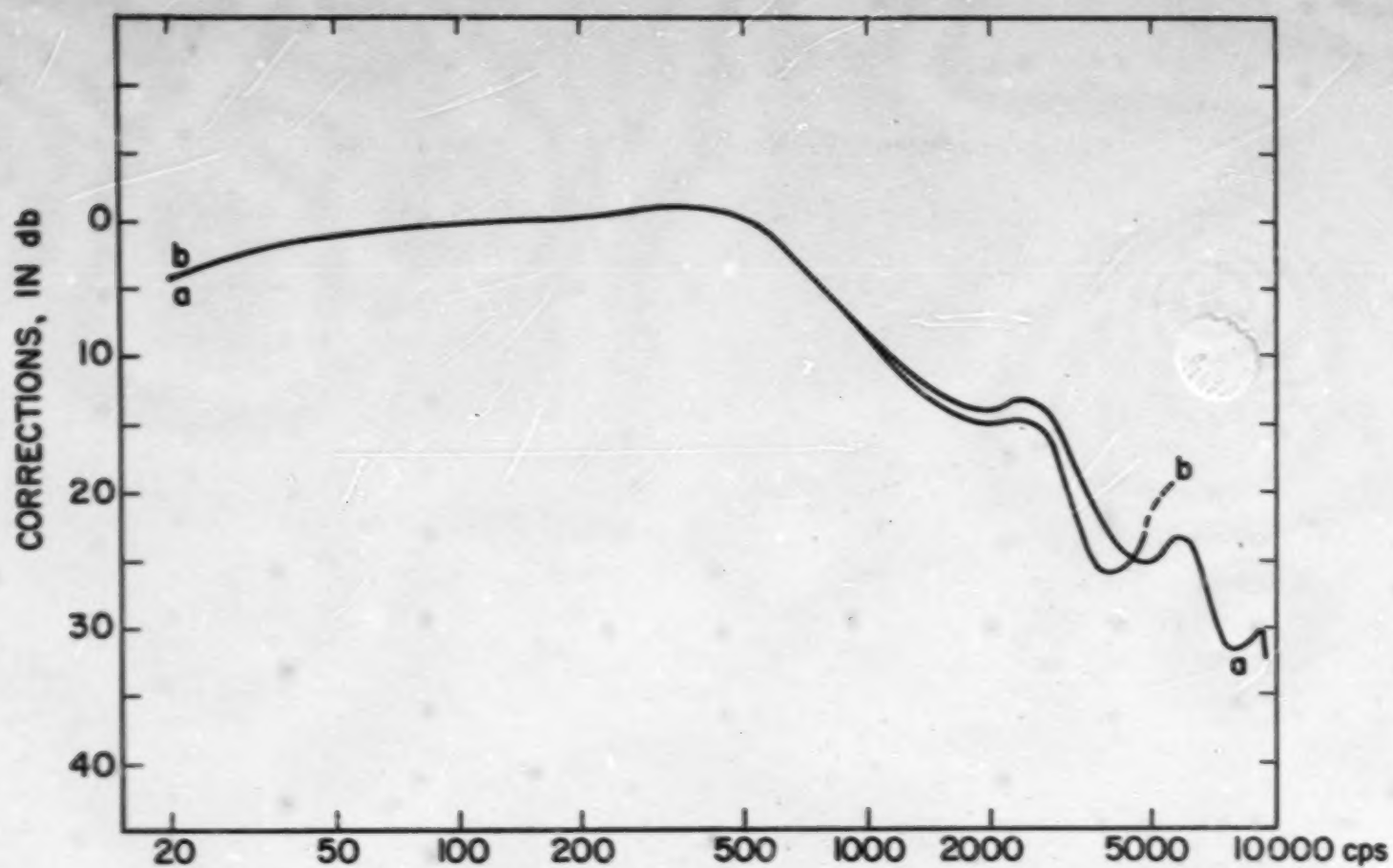


Fig. 1

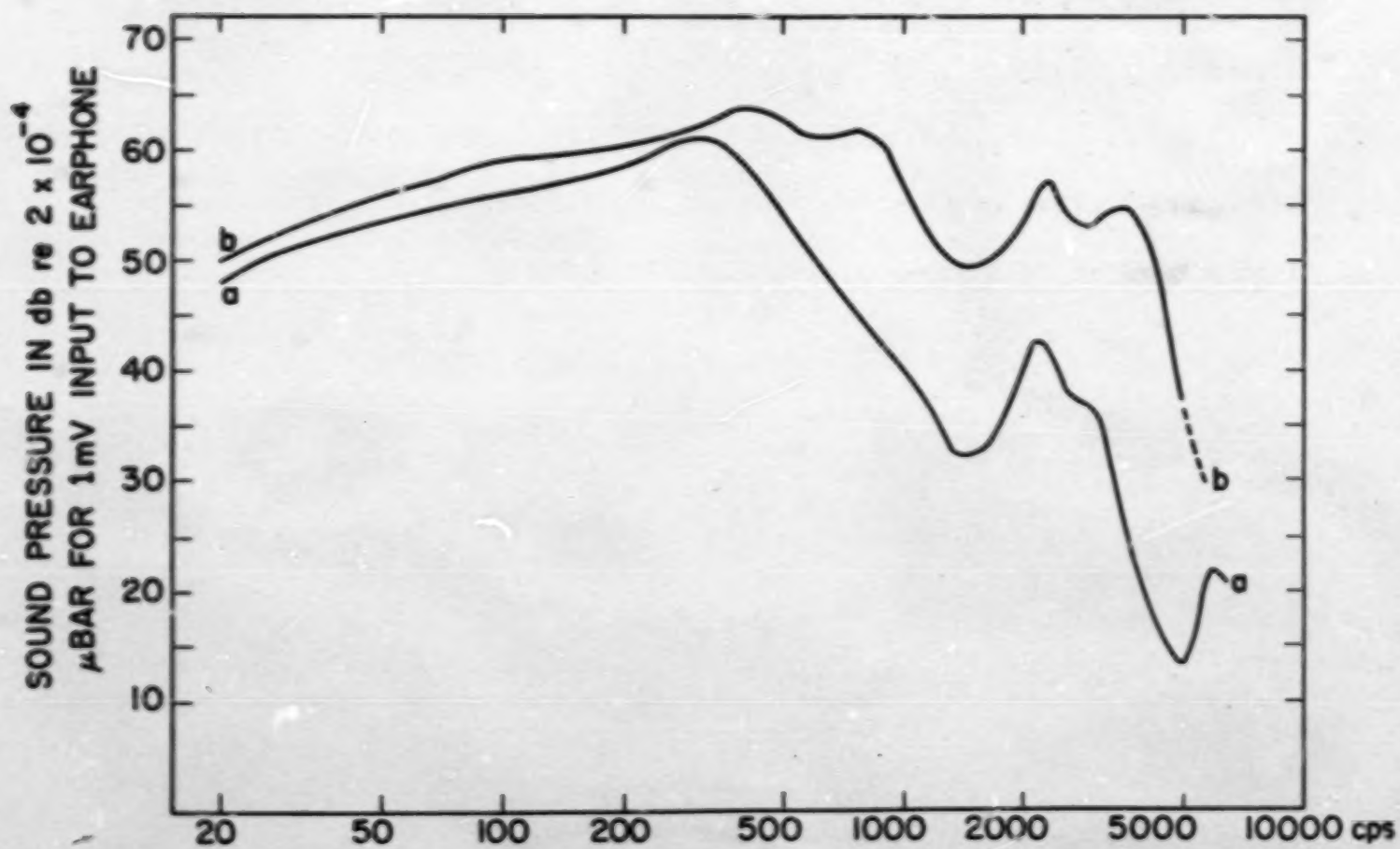


Fig. 2

APPENDIX

CURRENT STAFF

LIST OF PUBLICATIONS

Current Staff

Since May, 1963, the following personnel have been on the staff of the Behavior Analysis Laboratory.

H. L. Lane, Ph.D., Director
S. A. Ross, Mag. Art., Research Associate
D. V. Cross, A.M., Research Assistant
J. L. Kopp, A.M., Research Assistant
R. L. Buiten, B.S.E., Research Engineer
D. R. Brinkman, B.S.E., Administrative Assistant
W. C. Sheppard, Technical Assistant
D. M. Carlson, Programming Consultant
R. E. Skrocke, Research Technician
N. J. Kissock, Secretary

List of Publications

- Cross, D. V. & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. J. exp. anal. Behav., 1962, 5, 487-496.
- Kramer, E. F. Personality stereotypes in voice: a reconsideration of the data. J. Soc. Psychol. (in press).
- Kramer, E. F. The judgment of personal characteristics and emotions from nonverbal properties of speech. Psychol. Bull. (in press).
- Kramer, E. F. The elimination of verbal cues in judgments of emotion from voice. J. Abnorm. Soc. Psychol. (in press).
- Lane, H. L. Acquisition and transfer in auditory discrimination. Amer. J. Psychol. (accepted for publication).
- Lane, H. L. The autophonic scale of voice level for congenitally deaf subjects. J. exp. Psychol., 1963, 66, 328-331.
- Lane, H. L. Differential reinforcement of vocal duration. J. exp. anal. Behav. (in press).
- Lane, H. L. Foreign accent and speech distortion. J. Acoust. Soc. Amer., 1963, 35, 451-453.
- Lane, H. L. Programmed learning of a second language. In: Glaser, R. (Ed.) Teaching machines and programmed learning II: Data and directions. Dept. Audio Vis. Instruc., Natl. Educ. Assn., Washington, D.C. (in press).
- Lane, H. L. Psychophysical parameters of vowel perception. Psychol. Monographs, 1962, 761, Whole No. 563.
- Lane, H. L. Specifications for auditory discrimination learning in the language laboratory. Int. J. Amer. Ling., 1963, 29, 61-69.

- Lane, H. L., Hunting, J. R., Brethower, D. M. & Kopp, J. L. An effect of native language and musical training on vocal pitch matching. Percep. Mot. Skills (accepted for publication).
- Lane, H. L. & Kopp, J. L. The effects of response-dependent and independent reinforcement in extending stimulus control. Psychol. Rec. (in press).
- Lane, H. L. & Schneider, B. A. Methods for self-shaping echoic responding. Mod. Lang. J., 1963, 47, 154-160.
- Lane, H. L. & Schneider, B. A. Some discriminative properties of syntactic structures. J. Verbal Learn. Verbal Behav. (in press).
- Lane, H. L. & Shinkman, P. G. The relations among response rate, topography, and schedules of reinforcement: methods and findings in an analysis of the vocal operant. J. exp. anal. Behav., 1963, 6, 179-188.
- Schneider, B. A. & Lane, H. L. Ratio scales, category scales, and variability in the production of loudness and softness. J. Acoust. Soc. Amer. (in press).